C. IMPROVING HUMAN HEALTH AND PHYSICAL CAPABILITIES

THEME C SUMMARY

Panel: J. Bonadio, L. Cauller, B. Chance, P. Connolly, E. Garcia-Rill, R. Golledge, M. Heller, P.C. Johnson, K.A. Kang, A.P. Lee, R.R. Llinas, J.M. Loomis, V. Makarov, M.A.L. Nicolelis, L Parsons, A. Penz, A.T. Pope, J. Watson, G. Wolbring

The second NBIC theme is concerned with means to strengthen the physical or biological capabilities of individuals. The panel's work dovetailed with that of the first panel in the area of human cognition, especially the exciting and challenging field of brain performance. The *brain*, after all, is an organ of the human body and is the physical basis for that dynamic system of memory and cognition we call the *mind*. An extremely complex brain is the feature of human biology that distinguishes us from other animals, but all the other tissues and organs of the body are also essential to our existence and overall performance, and they thus deserve close scientific and technological attention.

The convergence of nano-bio-info-cogno technologies is bound to give us tremendous control over the well-being of the human body. In turn, it will change the way we think about health, disease, and how far we go to treat a patient. These new technologies will enable us to decipher the fundamental mechanisms of a living being, yet at the same time, they raise the fundamental questions of what life is and how human capability is defined. The panel gave highest priority to six technologies for the improvement of human health and capabilities in the next 10-20 years. In realizing these priorities, it will be essential to keep a "healthy" balance on human issues while seeking technological and social solutions.

1. Nano-Bio Processor

As the convergence of NBIC progresses, it will be imperative that the technology be focused on ways to help enhance human health and overall physical performance, be disseminated to a broad spectrum of the population, and be developed by a diverse group of scientists and engineers. One potential platform that will enable this would be a "bio-nano processor" for programming complex biological pathways on a chip that mimics responses of the human body and aides the development of corresponding treatments. An example would be the precise "decoration" of nanoparticles with a tailored dosage of biomolecules for the production of nanomedicines that target specific early biomarkers indicative of disease. The nanomedicine may be produced on one type of nano-bio processor and then tested on another that carries the relevant cellular mechanisms and resulting biomarker pathways. The nano-bio processor would parallel the microprocessor for electronics, such that the development of new processes, materials, and devices will not be limited to a handful of "nano specialists." With the advent of the nano-bio processor, knowledge from all fields (biologists, chemists, physicists, engineers, mathematicians) could be leveraged to enable advancements in a wide variety of applications that improve human health and enhance human capabilities.

2. Self-Monitoring of Physiological Well-Being and Dysfunction Using Nano Implant Devices

As the scales of nanofabrication and nanotransducers approach those of the critical biomolecular feature sizes, they give the technologist the toolset to probe and control biological functions at the most fundamental "life machinery" level. By the same token, this technology could profoundly affect the ways we manage our health.

One outcome of combining nanotechnology with biotechnology will be molecular prosthetics — nano components that can repair or replace defective cellular components such as ion channels or protein signaling receptors. Another result will be intracellular imaging, perhaps enabled by synthetic nanomaterials that can act as contrast agents to highlight early disease markers in routine screening. Through self-delivered nano-medical intervention, patients in the future will be able in the comfort of their homes to performed noninvasive treatments autonomously or under remote supervision by physicians.

Metabolic and anatomical monitoring will be able to give humans the capability to track the energy balance of intake and consumption. Monitoring of high-risk factors will be able to facilitate early diagnosis, when medical treatments can be most effective. Information systems designed to present medical data in ways that are intelligible to laypersons will allow anyone to monitor his or her health parameters. As a result of NBIC-enabled “wonder medicines,” there will be a need to develop technology and training modalities to make the patient an essential partner in the process of health monitoring and intervention.

As the population ages, more and more age-related diseases and deteriorating functions (e.g., hearing, memory, muscle strength, and sight) will be prevalent; an obvious example is Alzheimer’s disease. Some of these dysfunctions are due to molecular changes over time, and some are due to the natural decay of bodily functions. NBIC will provide ways to slow down the aging process or even reverse it.

3. Nano-Medical Research and Intervention Monitoring and Robotics

The convergence of nano-bio-info-cogno technologies will enhance the toolset for medical research and allow medical intervention and monitoring through multifunctional nanorobots. For example, a nano brain surveillance camera could be developed. Imaging tools will be enhanced by nanomarkers as anchor points for hierarchical pinpointing in the brain. A range of nano-enabled unobtrusive tools will facilitate research on cognitive activities of the brain.

Nano-enabled unobtrusive tools will be invaluable for medical intervention, for example, nanorobots accomplishing entirely new kinds of surgery or carrying out traditional surgeries far less invasively than does a surgeon’s scalpel. Technological convergence will also enhance post-surgery recovery. Although open surgical procedures will probably be reduced in numbers, the need for them will not be eliminated. Each procedure induces different side effects and risk factors. For instance, open-heart surgery increases the risk for stroke several days after the operation. NBIC technologies could enable devices that monitor these risk factors and immediately notify the physician at the first indication of a precursor to the onset of post-surgery traumas.

4. Multimodalities for Visual- and Hearing-Impaired

In the United States, there are 8 million blind people and 80 million who are visually impaired. The current paradigm of electronic communication is visual and conducted through the use of monitors and keyboards. It will be important for NBIC technologists to address the need for multimodal platforms to communicate with, motivate, and utilize this population group. Examples of different modes of communication include talking environments and 3-D touch screens to enable access to the Internet.

While convergent technologies will benefit disabled persons, they in turn will contribute greatly to the development of the technology, thereby benefiting all people. In recognition of this fact, disabled scientists and engineers should be included in research and design teams. As NBIC blurs the boundaries of normal and abnormal, ethical and unethical, it will be important to include disabled members and advocates on advisory committees at all levels. This will include the private sector, academia, government, and international committees.

5. Brain-to-Brain and Brain-to-Machine Interfaces

The communication among people and between people and machines or tools has not been fully realized because of the indirect interactions. The external tools need to be manipulated as an independent extension of one's body in order to achieve the desired goal. If machines and devices could be incorporated into the "neural space" as an extension of one's muscles or senses, they could lead to unprecedented augmentation in human sensory, motor, cognitive, and communication performance.

A major goal is to measure and simulate processes from the neuron level and then to develop interfaces to interact with the neural system. A visionary project by Llinas and Makarov proposes a nonintrusive retrievable cardiovascular approach to measure neuron and group-of-neuron activities, and on this basis, to develop two-way direct human communication and man-machine telepresence.

Another goal is to establish direct links between neuronal tissue and machines that would allow direct control of mechanical, electronic, and even virtual objects as if they were extensions of human bodies. Another visionary project by Nicolelis proposes electrophysiological methods to extract information about intentional brain processes and then translate the neural signals into models that are able to control external devices.

6. Virtual Environments

Nanotechnology will permit information technology to create realistic virtual environments and geographies. And biotechnology guided by cognitive science will produce interfaces that will allow humans to experience these environments intensely. Thus, the union of these technologies will transcend the biological limitations of human senses and create a new human relationship to the physical environment. It will be possible to simulate in humans the sensation of being at remote locations or at imaginary new buildings or facilities. This could be used for rapid design and testing of large projects, thereby saving the cost of errors. Other economically significant applications could be in the entertainment industry, and the tourist industry could use the technology to provide virtual samples of distant locations to prospective customers.

Applications of special relevance to improving health and enhancing human physical abilities include the use of virtual environments for education and interactive teaching. This will provide new ways for medical students to visualize, touch, enter, smell, and hear the human anatomy, physiological functions, and medical procedures, as if they were either the physician or a microscopic blood cell traveling through the body. Similarly, impaired users, ordinary people, athletic coaches, and a range of health-related professionals could train in these virtual environments.

Statements and Visions

Participants in the panel on human health and physical capabilities contributed statements and visions on a wide range of technological challenges and opportunities. Several contributors addressed life extension (P. Connolly); therapeutics at the cellular level (M.J. Heller, J. Bonadio), physiological level (A.T. Pope), and brain levels (B. Chance and K. A. Kang, E. Garcia-Rill, L. Cauller and A. Penz); as well as brain-machine interaction (R.R. Llinas and V. Makarov, M.A.L Nicolelis) and improving the quality of life of disabled people (G. Wolbring and R. Golledge).

Reference

Gazzaniga, M.S., ed. 1995. *The cognitive neurosciences.* Cambridge, MA: MIT Press.

STATEMENTS

NANOBIOTECHNOLOGY AND LIFE EXTENSION

Patricia Connolly, University of Strathclyde

This paper concentrates on only one of the complex debates emerging due to the convergence of nano-bio-info-cogno (NBIC) and the ability to improve human performance: that is, how nanobiotechnology will affect life extension. To deal with this in a comprehensive manner, the concept of life extension will be discussed, along with a brief presentation of the major obstacles that can be defined from our current knowledge in bioscience and medicine. It is proposed that a successful strategy for the convergence of NBIC disciplines in human terms will require a holistic approach and consideration of the full pathway from the human, down through organ, cell, and molecule, analyzing where NBIC can successfully intervene in this complex cascade. Some examples are given of areas where nanobiotechnology has had, or could have, impact in the problem areas of human well-being and quality of life as they are understood today.

Life Extension and Nanobiotechnology: Some Key Criteria

Nanobiotechnology for the purposes of this discussion is defined as the application of nanotechnology or nanobiology to a biological environment that involves device or material interactions with biological or biomolecular systems. To consider nanobiotechnology and life extension, it is important to first consider which social groups might be targeted by this approach and then to examine their medical and social requirements, highlighting where the NBIC convergence will have an effect. For example, the problems of the developed and developing world are quite different in terms of life extension. The problem of environmental damage and rising world pollution threatens the quality and length of life span of both groups. Table C.1 summarizes some of the major problems that must be addressed in extending life in developed and developing countries (WHO 1998a; WHO 1998b; WHO 2000; WHO 2001).

Table C.1
The Challenges to Life Extension in Developed and Developing Countries

Target Groups	Quality of Life Problems	Major Causes of Death and Disability
Developed Countries: Aging Populations only	Loss of strength and mobility Loss of mental sharpness / neurological disease Social isolation Poverty	Cardiovascular Disease Diabetes and its complications Inflammatory diseases including arthritis Cancer Neurological Disease or Impairment
Developing Countries: All age groups	Environmental, lack of safe water & sanitation Disease related loss of earnings Poverty	Malnutrition Infectious diseases Parasites Cardiovascular disease

Governments in the developed world, including the United Kingdom (UK Foresight Consultation Document 1999), have started to develop an awareness of the needs of the increasingly aged populations that they have and will have in the first half of this century. Major disease groups or medical conditions that are the major causes of death or disability in the aging populations of the developed countries of the world have been identified. For example, according to the World Health Organization (WHO 2000), in 1999 around 30 percent of deaths worldwide were caused by cardiovascular disease and 12 percent by cancer.

The problems of the developing world are quite different, and it might be argued that unless life extension in this environment is addressed by those who have the technology and wealth to do so, then the stability of developed societies worldwide will be affected. The medical problems of developing countries are widespread: many of these could be resolved by improvement in economic factors; however, some problems, such as parasitic infections, have eluded complete medical solutions. Toxoplasma infects 50 percent of the world population and leads to miscarriage, blindness, and mental retardation. The WHO (1998b) states that one child dies in the world every thirty seconds from malaria. There is much scope for improvement in the formulation of drugs, delivery modes, diagnostics, and effective vaccines for these and other diseases.

In addition, it is recognized that increasing levels of pollution with their consequent environmental changes drive aspects of both childhood and adult disease. The epidemiology of the disease patterns are being studied (WHO 2001), and nations are considering their role in reducing environmental emissions (EIA 1998). Nanobiotechnology may have a part to play here in land and water treatments through bioremediation strategies and in novel processes for industrial manufacture.

A Holistic Approach to Problem Definition

To effectively target emerging NBIC technologies, and in particular to make the most of the emerging field of nanobiotechnology, requires a strategic approach to identifying the problem areas in life extension. Biomedical problems currently exist on macro, micro, and nanoscales, and solutions to some apparently straightforward problems could enormously increase life expectancy and quality of life. A holistic approach would examine the key medical problems in the world's population that need to be solved to extend life, and at the same time, would consider the social environment in the aging population to ensure that quality of life and dignity are sustained after technological intervention.

A key element of this top-down approach is to consider the whole human being and not merely the immediate interface of nanobiotechnology with its target problem. The ability to view the needs in this area from a biomedical perspective that starts with the whole human and works down through organ and cellular levels to the molecular (nanoscale) level, will be an essential component of projects with successful outcomes in this field. There is little point in developing isolated, advanced technological systems or medical treatments to find that they solve one problem only to generate many others. For example, ingenious microdevices with nanoscale features that might patrol blood vessels or carry out tissue repairs have been suggested and designed (Moore 2001; Dario et al. 2000). However, there has been little detailed discussion or consideration at this stage regarding biocompatibility issues, particularly of the thrombogenicity (clot-forming potential) of these devices or of their immunogenicity (ability to stimulate an unwanted immune response). In this area, as in many others, there is a need for multidisciplinary teams to work together from the outset of projects to bring medicine and technology together. Ideally, these research teams would include clinicians, biomedical scientists, and engineers rather than being technologist-led projects that ignore much of the vast wealth of information we have already discovered about the human body through medical and biomedical research.

Accepting this need for biomedically informed project design also leads to the conclusion that understanding of the cell-molecule interface, in other words the micro-nanoscale interactions, will be a factor in the extended application of nanobiotechnology. To create a holistic approach to widespread and successful introduction of nanobiotechnologies in life extension will require interdisciplinary teams and exchange of information. Figure C.1 illustrates the possible levels of intervention and some of the emerging solutions where nanobiotechnology will have a role in repair or replacement of damaged elements.

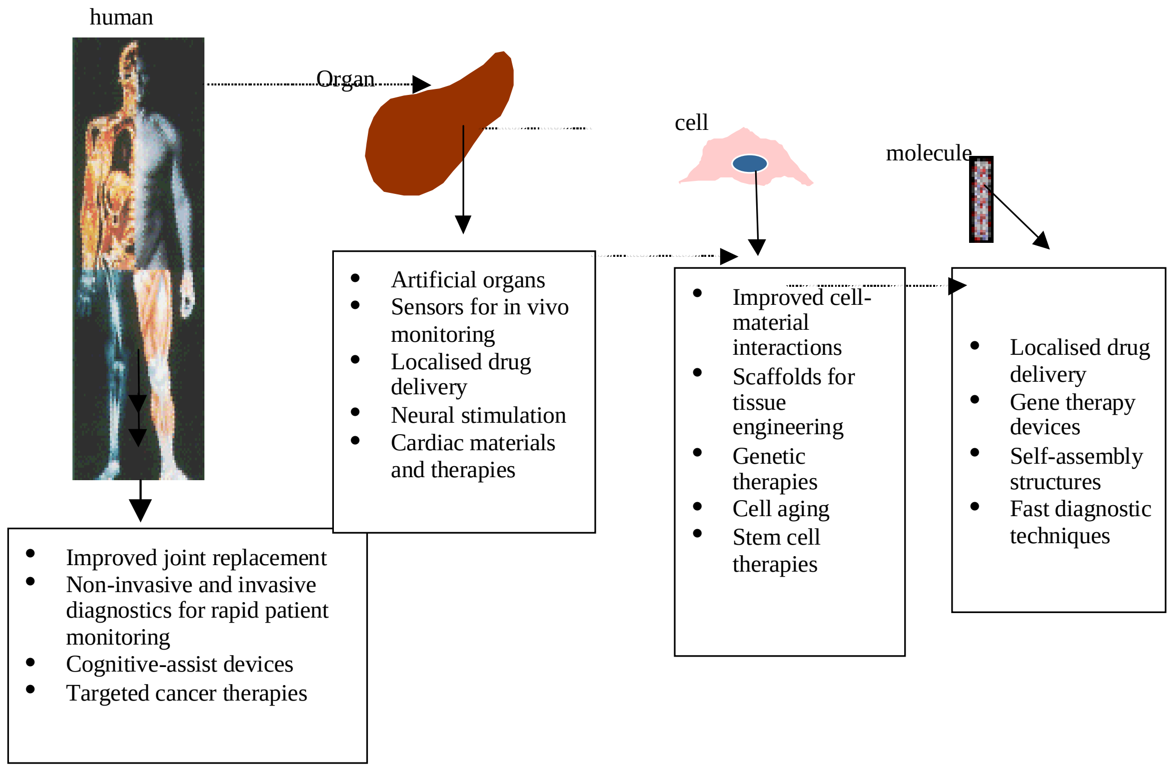


Figure C.1. Examples of levels for intervention of nanobiotechnology in human life extension.

The Need for a Holistic Approach: Some Specific Problems

As previously stated, there are a number of identified medical challenges that might benefit from intervention with nanobiotechnology. Many of these are long-term problems that have not been resolved by current technological or medical solutions. The following section is intended to briefly introduce some of these problems.

The Human-Materials Interface

Many of the disease conditions in the human body, and deaths during surgical intervention, can be traced to the body's in-built ability to react to foreign materials or wound sites through its inflammatory response. In normal disease or wounds, this ensures the proper activation of the immune response or of a clotting response from coagulation factors in blood. In extreme conditions or at chronic wound sites, the cascade reaction triggers a full inflammatory response that is harmful to tissue. In cardiovascular surgery, for example, reaction to physical intervention and surgical materials can lead to Systemic Inflammatory Response Syndrome (SIRS), and in a small percentage of cases, this will in turn lead to multiple organ failure and death (Khan, Spychal, and Pooni 1997).

The appearance of an inflammatory response following blood contact with a biomaterial can be readily measured in the molecular markers that are generated during the response, such as cytokines. (Weerasinghe and Taylor 1998). The reasons for the inflammatory response lie in molecular and cellular reactions at foreign surfaces. Nanobiotechnology could contribute to this field, both in terms of increasing the understanding of how the nanoscale events take place on particular materials and in terms of creating new, more biocompatible surfaces for use in surgery.

An extension of these problems is the continued reaction of the human body to any artificial implant, no matter how apparently inert the material. For the aging population, this has direct consequences as joints and tissues (such as heart valves) require replacement. Implanted replacement joints such as hip joints still suffer unacceptably high failure rates and shorter implantation life cycles than are ideal in an increasingly aged U.S. and European population. Hip implant rejection and loosening is caused by the interaction of cells with the coating or surface of the implant (Harris 1995). This can be modified, but not entirely halted, by drug interaction. The patient's cells react to both the materials and the micro and nanoscale surface features of the implant.

Nanobiotechnology has a place in the improvement of materials for surgery and implantation, both in the biological modification of surfaces to ensure that they do not degrade in use and in the study and manipulation of nanoscale topographies that directly influence cell movement and growth.

Neurological Disease

Both cellular decay and diseases such as Alzheimer's and Parkinson's contribute to loss of neural function, cognitive thought, and independence. In addition, events such as stroke leave many of the older population with impaired functions. It is here that implantable devices and cognitive science will have the greatest part to play in enhancing quality of extended life.

Microdevices for cell-electrode interfacing for both cardiac and neural cells have been available for *in vitro* applications for many years. There are few examples of implanted systems. Some micro-array type devices have been implanted, for example, in rudimentary artificial vision systems (Greenberg 2000). On a slightly larger scale, electrode systems have been implanted in the brain to provide electrical signal patterns that alleviate some symptoms of Parkinson's disease (Activa®, Medtronic, Inc., USA).

Much remains to be done in neurological device development, including devising smaller systems capable of withstanding long-term implantation. Investigation of the submicron (synaptic) interface to devices from neurons may be an important area for consideration in this field. In the longer term, it may be that some conditions will be alleviated by local electrode and drug-release systems, but how to keep these devices in place for years so that they remain biologically or electrically viable remains a difficult problem. There will be a need to develop sub-micron arrays of electrodes and chemo-arrays in devices designed to replace diseased tissue. If nanoscale electrode-cell interactions are expected to be important, then a fuller understanding of the cell-nanoelectrode interface will be required both *in vitro* and *in vivo*.

Cell replacement technologies are also being developed to address neural decay, and success with this type of approach, such as stem cells (see discussion of artificial organs and tissue engineering below), may remove the need for extensive device development. Cell placement and growth techniques may still, however, require device intervention and nanobiotechnology know-how.

Artificial Organs and Tissue Engineering

In the field of tissue repair and replacement, advances are being made in the creation of artificial organs and replacement tissue. In the case of artificial organs, many of the components of the organ will not be linked to the body's own regulatory systems (e.g., artificial heart pumps). In engineered tissue for repair or replacement of damaged tissue, control of tissue growth and tissue integration are critical and will require monitoring.

To provide sensitive feedback control to artificial organs either within or external to the body (such as the artificial liver), biosensor systems will be required, perhaps coupled to drug or metabolite delivery systems. This is an ongoing problem, since no long-term implantation systems based on biosensors have become commercially available, even with the application of microtechnology (Moore 2001; Dario et al. 2000) — although improvements have been made for subcutaneous glucose sensors in recent years (Pickup 1999). There is opportunity here for the use of nanobiotechnology to both provide the sensors for monitoring and adjusting organ performance and to aid localized drug or metabolite delivery to artificial organs. It may be possible to create biosensors for long-term implantation by trapping "factory cells" in gels within the sensor system, which would, in turn, synthesize any required renewable nanocomponents in the sensors, thus avoiding the current problems of sensor degradation over time.

Significant amounts of time, money, and research effort are being directed to the field of tissue engineering for skin, cartilage, bone, and heart tissue regeneration or repair, as well as for other types of tissue. Biopolymer scaffolds are the material of choice for the seeding of cells to grow replacement tissue. At the macro or fiber level, much is known about these scaffolds, but little time has been devoted to the nanoscale effects of topography or surface molecular treatments that could be influenced by nanobiotechnology. Nanovesicles that could be incorporated into tissue scaffold structures for slow release of chemoattractants could greatly improve tissue uptake or repair. One group has recently successfully exploited the idea of self-assembly of molecules, in this case, peptide-amphiphile molecules, to create biopolymer scaffolds with nanoscale features for bone repair (Hartgerink, Beniah, and Stupp 2001). This group's experiments show that a key constituent of bone, hydroxyapatite, can be made to grow and align in the same manner as bone *in vivo* using these scaffolds.

Stem cell research promises to open up new possibilities for harvesting cells that can be transformed *in situ* into different tissue types for repair or regeneration of damaged tissue. This may require extensive technological intervention both for harvesting cells and in delivering cells for therapy.

Genetic Techniques

The explosion in the field of genetics has led to the availability of a range of diagnostic tests for predisposition to illnesses, including cancer, although final expression of many illnesses may have strong environmental factors that must be taken into account. Together with the possibility of gene therapy for specific diseases, this offers new hope of life extension to many people. For example, hereditary lung conditions such as cystic fibrosis are being targeted by gene therapy to replace missing or deficient genes (Douglas and Curiel 1998). Study of how cells age is being taken up by many research groups and, again, offers hope for many potential victims of cancer and degenerative disease. Nevertheless, any widespread genetic intervention in disease is still some way off. To quote one recent review paper, "Ideally, gene therapy should be efficient, cell-specific, and safe (Hu and Pathak 2000). One of the challenges of gene therapy is the efficient delivery of genes to target cells. Although the nucleic acids containing the genes can be generated in the laboratory with relative ease, the delivery of these materials into a specific set of cells in the body is far from simple." It is perhaps here in the design and development of efficient delivery devices and systems that nanobiotechnology will play its biggest role in gene therapy.

Drug Delivery

There are still many opportunities for nanobiotechnology in the field of drug delivery, particularly in delivery of those drugs unsuitable for the gastrointestinal system. Skin and lungs have become favorite alternative routes for drug delivery, with nanovesicles and microcrystals as popular drug carriers (Langer 1999). Cancer treatment has yet to fully benefit from the targeted delivery to tumors of drugs in microdevices with local nanoscale interactions. Likewise, cancer monitoring and surgery would benefit enormously from miniaturized sensor or other diagnostics systems that could be used in the pre-, peri-, and postoperative environment.

The Prospects for Life Extension

Any quantitative discussion on the prospects for life extension through nanobiotechnology intervention in disease must be purely hypothetical at this stage. However, speculating across the human-organ-cell-molecule model may give some idea of the possible times to application of some of the approaches under development. Table C.2 summarizes what is a very personal view of the likely outcome of convergence in NBIC.

Table C.2
Some Potential Gains in Life Extension from NBIC convergence

Level of Intervention	Key Advance	Timescale	Life Extension
Human	Noninvasive diagnostics	5-10 years	Lifesaving for some conditions
	Cognitive assist devices	15-20 years	Higher quality of life for several years
	Targeted cancer therapies	5-10 years	Reduction in cancer deaths by up to 30%
Organ	Artificial heart	0-5 years	2-3 years awaiting transplant
	Neural stimulation or cell function replacement	5-20 years	10-20 years extra if successful for neurodegenerative patients
Cell	Improved cell-materials interactions	0-15 years	Lowering of death rates on invasive surgery by 10% and extending life of surgical implants to patient's lifetime
	Genetic therapies	30 years	Gains in the fight against cancer and hereditary diseases
	Stem cells	5-10 years	Tissue / brain repair Life extension of 10-20 years
Molecule	Localized drug delivery	0-10 years	Extending life through efficient drug targeting
	Genetic interventions	0-30 years	Life extension by targeting cell changes and aging in the fight against disease Likely to be a very complex environment to successfully manipulate

Visions for the Future

Loss of mobility and therefore independence is critical in the onset of decay and isolation for many older people, and one area in the developed world where people are very dependent for mobility is in the use of a car. Confidence and cognizance decline for many people as they age; in the car of the future there is the possibility to see the true convergence of NBIC in extending independence and warding off part of the decline in the older person. Higher-speed, higher-density computers and effective sensors driven by nanotechnology may combine with on-board artificial intelligence in the car, helping the driver plan routes and avoid hazards and difficult traffic situations. Nanobiotechnology may also be present in on-board minimally invasive biosensors to monitor the driver's health, both in terms of physical stress and physiological condition, to be fed back to the car's computer. In a further interpretation, since the possibility of implanted devices to stimulate or improve cognizance are emerging, the driver may be also benefit from neuronal stimulation designed to keep him or her alert and performing optimally during the trip.

The convergence of NBIC in the field of life extension will lead to implanted devices such as sensors and drug delivery systems being developed to replace or monitor body function. Implanted devices, whether macro or micro in scale, present a problem today in terms of biocompatibility. Implantation of a heart valve in a patient means that a drug regime for anti-coagulation is mandatory — usually through administration of warfarin. Since inflammatory response and immunogenic response take place *in vivo,* many of the devices being discussed and designed today to improve human performance incorporating nanotechnology will not be implantable because of biocompatibility issues. A further complication will be how to keep a nanodevice biologically or electronically active (or both) during sustained periods of operation *in vivo*. Sustained exposure to physiological fluid, with its high salt and water content, destroys most electronic devices. Likewise, devices that emit biological molecules or are coated with biological molecules to ensure initial biocompatibility must have their biological components renewed or be destined to become nonfunctional some time after implantation. Little attention is being given to these problems, which may prove major stumbling blocks in the next 10 to 30 years to the successful application of nanotechnology in a range of medical conditions.

A "holistic human project" could bring together the best research clinicians, biomedical engineers, and biomedical scientists to discuss the main life-shortening diseases and conditions and current progress or problems in their treatment or eradication. Together with the nanotechnologists, areas where conventional medicine has not been successful could be identified as strategic targets for nanobiotechnology. Specific project calls could follow in these areas, with the condition that the applicants' teams must show sufficient interdisciplinary interaction to provide a comprehensive understanding of the nature of the problem. The opportunities are immense, but the resources available are not unlimited, and only strategic planning for project groups and project themes will realize the maximum benefit for biomedicine and society.

References

Dario, P., M.C. Carozza, A. Benvenuto, A. Menciassi. 2000. Micro-systems in biomedical applications. *J. Micromech. Microeng.* 10:235-244.

Douglas, J.T., and D.T. Curiel. 1998. Gene therapy for inherited, inflammatory and infectious diseases of the lung. *Medscape Pulmonary Medicine* 2, 3.

EIA (Energy Information Administration, U.S. Dept. of Energy). 1998. *Impacts of the Kyoto Protocol on U.S. energy markets and economic activity*. Report No. SR/OIAF/98-03.

Greenberg, R.J. 2000. Visual prostheses: A review. *Neuromodulation*, 3(3):161-165.

Harris, W.H. 1995. The problem is osteolysis. *Clinical Orthopaedics and Related Research* 311: 46-53.

Hartgerink, J.D., E. Beniah, and S.I. Stupp. 2001. Self-assembly and mineralization of peptide-amphiphile nanofibers. *Science* 294: 1684-1688 (November).

Hu, W-S., and V.K. Pathak. 2000. Design of retroviral vectors and helper cells for gene therapy. *Pharmacological Reviews* 52: 493-511.

Khan, Z.P., R.T. Spychal, and J.S. Pooni. 1997. The high-risk surgical patient. *Surgical Technology International* 9: 153-166 (Universal Medical Press).

Langer, R. 1999. Selected advances in drug delivery and tissue engineering. *J. of Controlled Release*, 62: 7-11.

Moore, A. 2001. *Brave small world*. EMBO Reports, 2(2): 86-89. (European Molecular Biology Organisation, Oxford University Press).

Pickup, J. 1999. Technological advances in diabetes care. Wellcome News Supplement Q3(S).

UK Foresight Consultation Document. 1999. *The aging population*. (http://www.dti.gov.uk).

Weerasinghe, A., and K.M. Taylor. 1998. The platelet in cardiopulmonary bypass. *Ann. Thorac. Surg*, 66:2145-52.

WHO (World Health Organization). 2000. *1997-1999 World Health Statistics Annual*. Geneva: World Health Organization.

WHO. 1998a. *The World Health Report 1998, Life in the Twenty-first Century: A Vision for All* 1998, ISBN 92 4 156189 0

WHO. 1998b. *WHO Fact Sheet No. 94, Malaria*. Geneva: World Health Organization.

WHO. 2001. *Methodology for assessment of environmental burden of disease.* ISEE session on environmental burden of disease. Report of a WHO Consultation (WHO/SDE/WSH/00.7). Geneva: World Health Organization.

THE NANO-BIO CONNECTION AND ITS IMPLICATION FOR HUMAN PERFORMANCE

Michael J. Heller, University of California San Diego

Many aspects of nanotechnology will lead to significant improvements in human performance; however, the nano-bio area will be particularly important and relevant to such improvements. Technological advancements in the past decade have been nothing short of phenomenal. These advancements have led to an increasingly better understanding of human biology. We can expect that the new advancements in the nano-bio area will not just lead to a better understanding of human biology, but will also provide a new dimension and capability to affect human biology. The fact we are having this workshop and all know its true importance and underlying implications speaks for itself.

Individualized Treatment for Human Development

How nano-bio technologies will be applied in the most beneficial ways is dependent on the underlying basis for human performance. It is very likely that most of the underlying basis is genetic in origin (Wexler 1992; Ridley 2000). While this may still be widely debated and resisted for other reasons, it will (when proven) have profound implications, and it certainly needs to be considered in any planning on new technology application in human biology. The following is an example, which will hopefully not trivialize the issue.

Many individuals greatly enjoy a variety of sporting activities. However, a vast majority of individuals who do any of these sporting activities cannot approach the capabilities of a professional player, even with all the best new technology, instruction, and personal motivation. While some might feel this unfair, most people accept it and keep it in perspective. After all, people in general usually have something they do well, even if they never develop the desired trait. Not only is this true for athletic capabilities, but this is widely observed for other capabilities such as talent in art or music. Until recently, these perceptions were not based on any real scientific evidence. Now, with the first phase of the human genome project complete and a new geomics revolution occurring, good evidence is appearing that many human performance traits do indeed have a genetic basis.

This may also hold true for human behavior (Chorney et al. 1998; Dubnau and Tully 1998). Just a few years ago psychiatrists and psychologists would have doubted the genetic basis for many of the important mental illnesses. Today, there are few diseases left that are not known to be directly or indirectly genetically based (Kamboh 1995; Corder et al. 1994). Even infectious diseases are not really an exception to this premise, as there are always individuals who have a positive genetic component that provides varying degrees of resistance to the infection (Hill 1996).

A particularly relevant example of the importance of understanding the true basis of “cause and effect” in determining technological strategy now comes from the pharmaceutical industry. The new area of phamacogenomics is now proving for one drug after another that so-called drug toxicity is really based upon individual genetic polymorphisms. Usually, for any given drug, there are always a small number of individuals for whom that drug is toxic or less effective. As the genes and pathways for drug metabolism are better understood, this drug toxicity is usually found to correlate in some fashion with single nucleotide polymorphisms (point mutations) in the affected individuals. Not too long ago, most drug companies were investing huge amounts of money looking for “safe” drugs. Today, most accept or will soon accept the fact that patient stratification (via either genotyping or phenotyping) will be necessary to determine drug toxicity.

This represents a key example of how important it is to properly identify cause and effect in relation to technology development. The pharmaceutical industry spends enormous amounts of money developing new drugs, and many potentially useful drugs are being delayed or not used because they have serious toxicity for a small number of individuals. This also presents a view of how genetic determination is misunderstood. If we were to look at just a single drug, genetic testing of potential drug recipients would seem totally unfair and appear that genetic testing is being used to exclude some individuals from a potential benefit — even though some individuals truly don’t benefit from that particular drug. However, at least in the area of therapeutics, we do not have to look at too many drugs until we find that, in general, the vast majority of humans will always have one or two important drugs that are not beneficial or are harmful to them. The lesson here is that it does not do a lot of good to pump enormous amounts of money into developing technology for new drug discovery without patient stratification — and this is genetics.

We should probably expect the same scenario to develop for human performance, and also, whether we like it or not, for human behavior.

Thus, now is really the time for scientists to put this issue into proper perspective. The misconception and fears about genetic determination are so misguided that we are delaying technology that can actually help improve existence for everyone. In medical diagnostic areas, we accept without any reservations tests and assays that try to determine if we have a disease or the state of that disease. However, many people view with great concern genetic testing that is more direct and provides earlier detection. There are most certainly very important ethical issues relevant to the genetic determination. But even these are in some sense clouded by misconceptions, due to past behavior by groups who misunderstood the real meaning of genetic determination and/or intended to misuse it. It is time to correct this and gain the full benefits of our technology for everyone.

Tentative Plan for Understanding Genotype and Performance

We should start with the premise that almost every (physical) performance trait will be related to some distinct group of genotypes. (Genotypes from outside the group can also influence the trait, but this does not change the basic premise). This group of related genotypes will usually present itself in the general population as most individuals having average performance, some individuals having below-average performance, and another group of individuals having above-average performance. If we were to take “*running*” as an example, we can already begin to scientifically relate this trait to genetic polymorphisms in muscle tissue as well as other physiological characteristics. Even though we will ultimately identify the related group of genotypes that can accurately predict the performance level for any given physical trait, several problems do exist. The first problem is that there is considerable complexity in how different traits combine to affect “overall” performance. The second problem is to determine how these combinations of traits influence overall performance under different environmental challenges or stresses.

The goals for an initial plan to evaluate genotype and performance are listed below:

i) Begin to correlate physical (and related behavioral) performance characteristics with the genotypes and polymorphisms that are rapidly emerging from the human genome project. This would not be much different than what pharmaceutical companies are doing related to patient stratification for drug toxicity effects.

ii) Begin to model how combinations of traits influence overall performance. Then separate the groups of directly related genotypes from those that indirectly influence the trait.

iii) Begin to model and understand how a higher performance trait (or traits) that provide(s) an advantage under one set of environmental conditions and/or challenges, is not an advantage or is even a disadvantage under another set of environmental conditions and/or challenges.

This third point is probably the most difficult to deal with, because it leads to diversionary semantic and philosophical questions as to whether biology (genetics) or environment is in control, and what is cause and what is effect. These questions will be put into better perspective using examples of genetic disease in the human population (Jorder et al. 2000) and examples of how particular "types" of stress relate to heart disease (Ridley 2000; Marmot et al. 1991).

References

Hill, A.V.S. 1996. Genetics of infectious disease resistance. Opinion in *Genetics and Development* 6: 348-53.

Chorney, M.J., et al. 1998. A quantitative trait locus associated with cognitive ability in children. *Psychological Science* 9: 1-8.

Corder, E. H. et al. 1994. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer's disease. Nature Genetics 7: 180-84

Dubnau, J. and T. Tully. 1998. Gene discovery in drosophilia: New insights for learning and memory. Annual Review of Neuroscience 21: 407-44.

Jorde, L.B., J.C. Carey, M.J. Bamshed, and R.L. White. 2000. Medical genetics. 2nd ed. St. Louis, MO: Mosby.

Kamboh, M.I. 1995. Apolipoprotein E polymorphisms and susceptibility to Alzheimer's disease. Human Biology 67: 195-215.

Marmot, M.G., et al., 1991. Health inequalities among British civil servants: The Whitehall II study. Lancet 337: 1387-93.

Ridley, M. 2000. Genome: The autobiography of a species in 23 chapters. New York: Prennial/Harper Collins.

Wexler, N. 1992. Clairvoyance and caution: Repercussions from the Human Genome Project. In *The code of codes.* D. Kevles and L. Hood, eds. Cambridge, MA: Harvard University Press.

GENE THERAPY: REINVENTING THE WHEEL OR USEFUL ADJUNCT TO EXISTING PARADIGMS?

Jeffrey Bonadio, University of Washington

The availability of the human genome sequence should (a) improve our understanding of disease processes, (b) improve diagnostic testing for disease-susceptibility genes, and (c) allow for individually tailored treatments for common diseases. However, recent analyses suggest that the abundance of anticipated drug targets (yielded by the genome data) will acutely increase pharmaceutical R&D costs, straining the financial outlook of some companies. Therefore, to stay competitive, companies must couple a threshold infrastructure investment with more cost-effective validation/development technology. However, no such technology currently exists.

This paper discusses the potential advantages and disadvantages of gene therapy as a validation/delivery platform for the genomics era. Gene therapy is the use of recombinant DNA as a biologic substance for therapeutic purposes. Although significant technological hurdles exist, for certain drug targets the potential for gene therapy as a validation/delivery platform are enormous. Thus, one may see

- direct, efficient transitions from database query to preclinical validation to lead drug candidate development

- significant improvements in the patient care pathway of important common diseases such as cancer, diabetes, and osteoporosis; these improvements would be expected to yield improved compliance and significantly better control of disease manifestations

The vision is that in 10 to 15 years, the U.S. private sector will have a drug discovery and development pathway that is significantly more cost-effective than what exists now and therefore is capable of taking full advantage of the promise of the human genome database. If this vision is realized, one can easily imagine that the process of transferring advances in drug development from the developed world to the undeveloped world will be significantly enhanced.

To traverse the technological hurdles associated with this vision, an interdisciplinary spirit will be required to advance our knowledge base in basic science and drug development, e.g., geneticists will (again) need to talk to physicists, physiologists to chemists, and cell biologists to engineers.

Drug Development Trends: Personalized Medicines

Human health is determined by the satisfaction of basic needs such as food and the avoidance of serious hazards such as trauma, environmental change, or economic disruption. In the world today, we find examples of almost all forms of social organization that have ever existed, including communities of hunter-gatherers, nomadic pastoralists, and primitive agriculturalists; unhygienic, large cities in the third world; and modern, large cities of the developed world. This variation in living conditions is associated with differing patterns of human disease around the globe (McKeown 1988) as well as with patterns that shift in a dynamic manner, creating a rather large and varied number of therapeutic targets for the pharmaceutical industry to consider.

In contrast to the dynamic and varied patterns of human disease worldwide, the pharmaceutical industry has a long history of pursuing only those limited number of human proteins (G-protein coupled receptors, ion channels, nuclear hormone receptors, proteases, kinases, integrins, and DNA processing enzymes) that make the best drug targets (Wilson et al. 2001). Even so, a high percentage of drug candidates never reach the market because adverse reactions develop in a significant percentage of individuals, while many approved drugs are effective for only a fraction of the population in which they are prescribed. This variation in drug response depends on many factors, including gender, age, genetic background, lifestyle, living conditions, and co-morbidity.

Since the 1950s, pharmacogenetic studies have systematically identified allelic variants at genetic loci for relevant drug-metabolizing enzymes and drug targets (Evans and Relling 1999). These studies suggest that genetic tests may predict an individual's response to specific drugs and thereby allow medicines to be personalized to specific genetic backgrounds. For some drugs, the geographic distribution of allelic variants helps explain the differences in drug response across populations. The frequency of genetic polymorphisms in drug-metabolizing enzymes, which contribute significantly to phenotype, may vary among populations by as much as twelve-fold. For example, between 5 percent and 10 percent of Europeans, but only 1 percent of Japanese, have loss-of-function variants at *CYP2D6* (debrisoquine oxidation) that affect the metabolism of commonly used agents such as beta-blockers, codeine, and tricyclic antidepressants. Polymorphisms in drug-metabolizing enzymes can lead to acute toxic responses, unwanted drug–drug interactions, and therapeutic failure from augmented drug metabolism (Meyer and Zanger 1997). Therefore, one approach to drug development in the future may be to test candidate formulations in populations that are genetically homogenous for certain key genetic markers. Still, specific research challenges remain as to the most appropriate way to catalog human genetic variation and relate the inferred genetic structure to the drug response.

Impact of Genome Analysis Technology

The preceding fifty years have been a time of rapid and profound technological change. The elucidation of the genetic flow of biological information (i.e., information flow from DNA to RNA to protein) has provided a basis for the development of recombinant DNA technology; the rise of molecular cell biology; the advent of intellectual property in biology and medicine); the development of the biotechnology industry; the development of transgenic technologies (including human gene therapy); the elucidation of the modern definition of stem cells; and the advent of cloning technology. Arguably, the defining technological event of the last few years has been the development and large-scale implementation of tools for the global analysis of genomes. Less than a decade ago, it was relatively uncommon to have full-length cDNAs at hand for experimental purposes. Within a decade, it may be commonplace to freely access the atomic structure of proteins, often in the context of their molecular partners. We have entered a new era of life science discovery research in which structure-function relationships form the basis of our understanding of cellular physiology and pathology (Ideker, Galitski, and Hood 2001).

We have also entered a new era of pharmaceutical discovery in which structure-function relationships underlie the search for new therapies (Dry, McCarthy, and Harris 2001). Thus,

- *We still do not know how the transcription machinery regulates gene expression* (Strausberg and Riggins n.d.), despite the fact that the scientific literature richly describes the presence and functional significance of alternatively processed human transcripts — as derived from different transcription initiation sites, alternative exon splicing, and multiple polyadenylation sites. Therefore, genome sequences must be annotated and additional databases of information must be developed.

  Large-scale analysis of gene expression originates from the expressed sequence tag (EST) concept. In the EST approach, a unique identifier is assigned to each cDNA in a library. Sequence tags of more than 700 nucleotides are now common, and the EST approach has been aided by formation of the IMAGE consortium, an academic-industrial partnership designed to distribute clones. The Merck Gene Index and the Cancer Genome Anatomy Project produced many of the human clones distributed through the IMAGE consortium (http://image.llnl.gov/).

  Imaginative new strategies complement the traditional EST approach. One of these, "serial analysis of gene expression" (Velculescu, Vogelstein, and Kinzler 2000), produces sequence tags (usually 14-nucleotides in length) located near defined restriction sites in cDNA. One advantage of this method is that each transcript has a unique tag, thereby facilitating transcript quantification. Tags are concatemerized, such that 30 or more gene tags can be read from a single sequencing lane, which also facilitates the effort to catalog genes. The Cancer Genome Anatomy Project, working together with the National Center for Biotechnology Information, has generated a SAGE database, SAGEmap, that includes over 4,000,000 gene tags. To proceed effectively with transcriptome efforts, there has been a significant shift in emphasis toward the sequencing of complete human transcripts.

In this regard, in 1999 the National Institutes of Health announced the Mammalian Gene Collection Project (http://mgc.nci.nih.gov), which aims to identify and sequence human and mouse full-length cDNAs. To date, that project has produced over 5,000 human sequences (deposited in GenBank). The German Genome Project recently completed full-ORF human cDNA sequences derived from 1,500 human genes.

- *Functional genomics may provide a mechanism to understand how proteins collaborate in an integrated, regulated, adaptive manner*. Multiple technologies support the field of proteomics, including genomics, microarrays, new mass spectrometry approaches, global two-hybrid techniques, and innovative computational tools and methods (Fields 2001). Protein localization within cells is now feasible at a genomic level. For example, thousands of yeast strains were generated recently in which more than 2000 *S. cerevisiae* genes were marked by transposon tagging (Ross-Macdonald et al. 1999). Indirect immunofluorescence was used to determine the subcellular localization for over 1,300 of the tagged proteins.

  Increasingly, proteomic strategies afford the opportunity for quantitative analysis of the cellular response to environmental change. Advances in direct analysis by mass spectrometry of peptide mixtures generated by the digestion of complex protein samples have lead to an escalating number of protein identifications in one experiment. These and other advances suggest that human tissues one day may be evaluated this way to advance our understanding of disease etiology and pathogenesis.

  Finally, protein expression and purification technologies will continue to improve, and procedures that make use of protein arrays will become commonplace. Potential applications include revealing interactions among proteins and between proteins and small molecules (drugs) or other ligands. The promise of this approach was suggested by the recent demonstration of proteins in nanoliter droplets immobilized by covalent attachment to glass slides: more than 10,000 samples could be spotted and assayed per slide with this technique (MacBeath and Schreiber 2001).

  A shift from genomics to proteomics is likely to be complicated, because single genetic loci may yield multiple polypeptides; proteins may change conformation in order to carry our a particular function; protein levels often do not reflect mRNA levels; proteins may undergo post-translational modification and proteolysis; and the presence of an open reading frame does not guarantee the existence of a protein. Proteins may also adjust their stability, change locations in the cell, and swap binding partners.

  Finally, protein function may depend on context, i.e., the function of an individual protein may be determined by the entire set of proteins operating in a microenvironment at a particular point in time — the concept of protein pleiotropism (Sporn 1999). When taken together, these considerations suggest that the proteome may be an order of magnitude more complex than the genome (Fields 2001; Hol 2000).

- *Structural genomics promises to capitalize upon numerous advances in cloning, protein expression, protein purification, characterization, crystallization, crystal drop inspection, crystal mounting, model building, and NMR spectra interpretation*, although high-throughput structure determination of drug candidates is not yet available (Russell and Eggleston 2000). With the potential to impact heavily on the design of new pharmaceuticals, structural genomics will take a place alongside high-throughput chemistry and screening as an integral platform approach underpinning modern drug discovery. Like the large-scale genomic sequencing projects that have been running for more than a decade, this will involve profound changes in thinking and approach.

Instead of developing a specific biological justification in advance of working on a protein, crystallographers and NMR spectroscopists can now consider the determination of structures for all proteins in an organism.

Bioinformatics will play several roles in structural genomics. Target selection involves database interrogation, sequence comparison, and fold recognition in order to aid selection of the best candidate proteins given a particular set of requirements, e.g., disease-associated genes, or those that are common to most organisms. Solved structures must be placed in an appropriate genomic context and annotated so that functional details may be predicted. Structural annotation may prove tricky, since large numbers of proteins of known structure but of unknown function have not previously been a major issue. Comparative modeling plays an essential role by providing structures for homologs of those determined experimentally, and efficient archiving of structural information is essential if the biological community is to make best use of all data. Given the biological and technological complexity associated with genome analysis technology, an interdisciplinary spirit will be essential to advance our knowledge base in basic science and drug development.

Drug Development in the Era of Genome Analysis: Applied Genomics

From SNP maps to individual drug response profiling, the human genome sequence should improve diagnostic testing for disease-susceptibility genes and lead to individually tailored treatment regimens for individuals with disease. Recent analyses (from both the public and private sector) suggest that the abundance of anticipated drug targets will dramatically increase pharmaceutical R&D costs. For example, it has been suggested that a threshold investment of $70-100 million will be required if companies are to profit from recent advances in bioinformatics. However, this investment may not yield a near-term return because current validation/development methods for drug targets are insufficiently robust to add value to R&D pipelines. Competitive considerations require companies to couple considerable infrastructure investment with cost-effective validation and/or development technology that has yet to be developed.

As described above, with advances in technology, the rational design and validation of new therapeutics increasingly will rely on the systematic interrogation of databases that contain genomic and proteomic information. One can imagine three pathways from database discovery to a validated product prototype, as shown in Figure C.2.

For *Pathway 1, rational small-molecule design*, the methods for developing a small-molecule prototype are well established in the pharmaceutical industry, which reduces risk. However, it is not clear that small-molecule drugs can be designed, as shown above: the notion currently is without precedent (with perhaps the exception of inhibitors of HIV protease and influenza neuraminidase), and therefore is best considered as an unproven hypothesis.

A major advantage for *Pathway 2, recombinant protein/peptide design*, is that small-molecule prototypes need not be designed and validated at all, which may significantly accelerate product development. However, therapeutic peptides and recombinant proteins are generally ineffective when administered orally, and alternative routes of administration are generally associated with challenges in terms of formulation, compliance, efficacy, and safety.

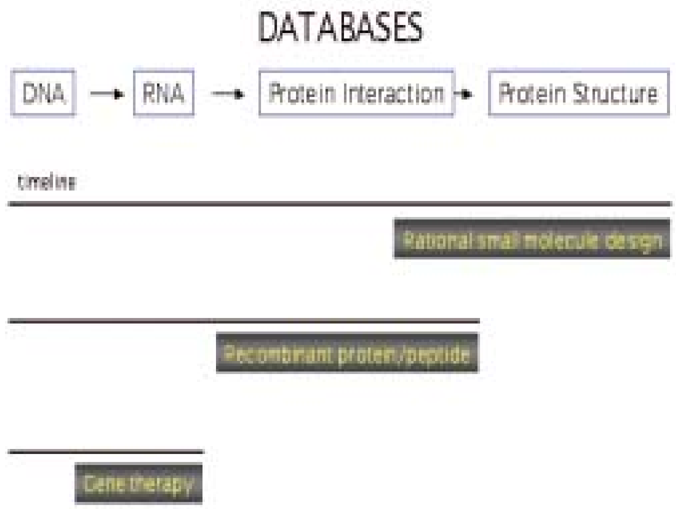


Figure C.2. Three pathways of drug discovery and development in the bioinformatics era.

A major advantage for *Pathway 3, gene therapy design*, is that one may proceed directly from database query to gene-based prototype — in theory, the shortest route to product validation and development. However, gene therapy is an early-stage technology, with known challenges in erms of efficacy and safety.

The Potential for Gene Therapy as a Validation / Delivery Platform

Gene therapy is the use of recombinant DNA as a biologic substance for therapeutic purposes (Bonadio 2000). Both viral and nonviral vectors have been employed. Nonviral vectors show many formulation and cost advantages, and they present a flexible chemistry. For example, the formulation of nonviral vectors with cationic agents results in nanometer-sized particles (synthetic polyplexes and lipoplexes) that show good efficiency (Felgner et al. 1997). Nonviral vectors have no theoretical sub-cloning limit, show a broad targeting specificity, transfect cells as episomes, and can be manufactured at scale relatively inexpensively. To enhance efficiency even further, one may use PEG to control surface properties of synthetic complexes, incorporate targeting moieties, use tissue-specific promoters, and incorporate fusogenic peptides and pH-responsive polymers.

On the other hand, the gain in gene-transfer efficiency associated with synthetic complexes must be balanced against the general lack of stability of polyplex and lipoplex vectors *in vivo* and the tendency of locally delivered cationic agents to cause tissue necrosis, which can be dramatic. Nonviral vectors are inefficient, and high doses may be required to achieve therapeutic effects. High-dose administration may be limited, however, by motifs in the vector backbone that stimulate the immune system (MacColl et al. 2001). While CpG-dependent immune stimulation is Th1-biased, SCID mice (Ballas, Rasmussen, and Krieg 1996) have shown increased levels of IFN- and IL-12 following plasmid-vector delivery (Klinman et al. 1996). Significantly, nonviral vector administration to animals has generated anti-DNA antibodies, leading to renal disease and premature death (Deng 1999). Relevant to the present application, Payette and colleagues (2001) recently showed that intramuscular delivery of a nonviral vector vaccine in mice led to destruction of antigen-expressing myocytes via a CTL-response.

Viruses are natural vectors for the transfer of recombinant DNA into cells. Recognition of this attribute has led to the design of engineered recombinant viral vectors for gene therapy. Viral vectors from retroviral, lentiviral, adenovirus, and herpes simplex species provide an important advantage in that they maximize gene transfer efficiency (Kay, Glorioso, and Naldini 2001). Viral genomes consist of genes and *cis*-acting gene regulatory sequences. Although overlap exists, most *cis*-acting sequences map outside viral coding sequences, and this spatial segregation is exploited in the design of recombinant viral vectors. Additionally, coding sequences work in *trans*, and viral genomes can be expressed by heterologous plasmids or be incorporated in the chromatin of producer cells to ensure stability and limit remobilization. Therefore, to generate vector particles, therapeutic genes and *cis*-acting sequences are first subcloned into separate plasmids, which are introduced into the same cell. Transfected cells produce replication-defective particles able to transduce target cells.

Viral vectors have inherent properties that affect suitability for specific gene therapy applications. A useful property of retroviral vectors, for example, is the ability to integrate efficiently into the chromatin of target cells. Disruption of the nuclear membrane is absolutely required for the pre-integration complex to gain access to chromatin (Roe et al. 1993), and productive transduction by retroviral vectors is strictly dependent on target cell mitosis (Miller, Adam, and Miller 1990). (Integration does not, however, guarantee stable expression of the transduced gene.) Because only a small fraction of muscle fibers pass through mitosis at any given time, this effectively prevents the use of regulated retroviral vectors (Rando and Blau 1994) in direct *in vivo* muscle gene therapy.

In contrast, replication-defective Ad vectors are attractive because they transduce post-mitotic cells very efficiently *in vivo* (Kozarsky and Wilson 1993). However, Ad vectors induce toxic immune responses that abrogate gene expression (Yang et al. 1995; Somia and Verma 2000). In a relevant example, Rivera et al. (1999) studied the feasibility of regulated Ad gene delivery after intramuscular injection in mice. The investigators employed an Ad vector cocktail encoding human growth hormone (hGH) under the control of transcriptional-switch technology. In initial experiments using immune-deficient mice, a single IP injection of rapamycin (5.0-mg/kg) resulted in a 100-fold increase in the plasma hGH level. Levels then diminished to baseline over the next 14-days. Similar induction profiles were noted after five subsequent injections (administered periodically over 6-months), and a direct relationship was observed between the peak hGH level and the amount of rapamycin administered (the i.v. dose range was 0.01- to 0.25 mg/kg). However, in immune-competent animals, peak levels of hGH were 50-fold lower, and no induction was observed after the first administration of rapamycin. These results were attributed to the destructive cellular and humoral immune responses to the Ad vector.

Experience with gene therapy suggests that this technology could serve as a broad validation and delivery platform for Pathway 3. To succeed, however, gene therapy must become a technology that more closely conforms to the current framework for drug development by pharmaceutical companies. Toward this end, gene therapy will need to be more easily managed by physician and patient; capable of producing therapeutic protein in a precise, dose-responsive, controllable manner; and formulated in a more simple, stable, and inexpensive manner. Ideally, a controllable gene-delivery system should feature low baseline transgene expression, a high induction ratio, and tight control by a small molecule drug. Indeed, it is difficult to imagine any gene-therapy (for any indication) that does not involve regulated therapeutic gene expression as a way to avoid toxicity and still respond to the evolving nature of disease.

Among a multiplicity of DNA vector alternatives, recombinant adeno-associated viral (rAAV) vectors (Monahan and Samulski 2000) represent an attractive choice for a validation and delivery platform. rAAV vector particles efficiently transduce both dividing and nondividing cells, and the rAAV genome persists as integrated tandem repeats in chromosomal DNA. (Upon co-infection with helper virus, AAV also transduces cells as an episome.) Elimination of AAV *rep* and *cap* coding sequences from rAAV prevents immune responses to viral gene products and the generation of wild-type helper virus (Hernandez et al. 1999; Xiao, Li, and Samulski 1996; Jooss et al. 1998). Transgene expression *in vivo* typically reaches a steady state after a gradual 2- to 10-week rise. Together, host chromosome integration and the absence of a cytotoxic T lymphocyte response provide a viable mechanism for long-term transgene expression, as demonstrated in skeletal muscle (Herzog et al. 1999; Malik et al. 2000; Ye et al. 1999; Herzog et al. 1997) and brain (Davidson et al. 2000) of immunocompetent animals, and in skeletal muscle of human subjects (Kay et al. 2000). Importantly, the ability to conduct experiments is supported by the availability of small-scale procedures that allow the facile manufacture of sterile rAAV preparations at titers of 10^{11}-10^{12} vector genomes/mL (Auricchio et al. 2001). Even more importantly, rAAV gene therapy is controllable, as demonstrated below.

One promising technology (Figure C.3) employs a heterologous transcription factor that selectively binds the transgene promoter and activates transcription in response to a cell-permeant controller molecule (e.g., Rivera et al. 1996; Magari et al. 1997; Pollock et al. 2000). Activation is achieved by reconstitution of a transcription factor complex that couples independently expressed protein chimeras (Brown et al. 1994; Standaert et al. 1990). One protein consists of a unique DNA-binding domain called ZFHD1, genetically fused to FKBP. The other protein chimera consists of the activation domain of the p65 subunit of NFκB, fused with the rapamycin-binding domain of FRAP, which is termed FRB. Packaging limits of rAAV require that the three components of the system be incorporated into two vectors, one vector that expresses both transcription factors from a single transcriptional unit, and a second vector containing the therapeutic gene driven by a promoter recognized by the ZFHD1 DNA-binding domain. Infection of permissive human cells with equal quantities of the two AAV vectors at a high multiplicity of infection has resulted in full *in vitro* reconstitution of the regulated system with at least a 100-fold induction after exposure to rapamycin. (Effectiveness may be dramatically increased [Mateson et al. 1999] when chimeric transcriptional activators are expressed as noncovalent tetrameric bundles.)

The feasibility of reconstituting the regulated system *in vivo* has also been determined. Skeletal muscle has been selected for local delivery because muscle is permissive for rAAV transduction, and because its component cells (muscle fibers) are long syncytia with extended nuclear domains that may be independently transduced with each vector. In one example (Ye et al. 1999), a controllable rAAV vector cocktail ($2x10^8$ infectious particles, with rAAV vectors at a 1:1 ratio) was injected into skeletal muscle of immune-competent mice. The administration of rapamycin resulted in 200-fold induction of erythropoietin in the plasma. Stable engraftment of this humanized system was achieved for 6 months, with similar results for at least 3 months in an immune-competent rhesus model.

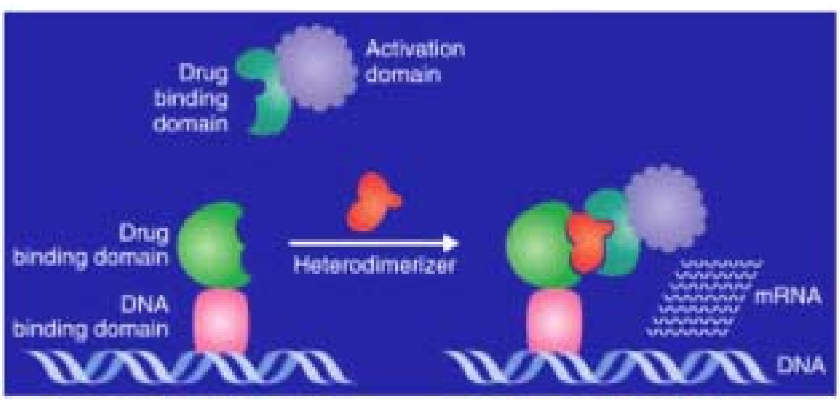


Figure C.3. Controlling gene expression using regulated transcription.

The "transcriptional-switch" technology (described above) features an induction-decay response for the therapeutic protein that occurs on a time-scale of days: transgene-encoded protein in blood typically peaks at about 24 hours and then decreases to background over 4 to 14 days. This kinetic profile probably reflects the "early-point" of transgene regulation as well as the many potentially rate-limiting steps after therapeutic gene delivery. These steps involve the pharmacokinetics and pharmacodynamics of rapamycin (Mahalati and Kahan 2001) as well as the dynamic processes of transgene transcription, therapeutic protein translation and secretion, and therapeutic protein bioavailability. Such prolonged kinetics may be appropriate for certain proteins (e.g., erythropoietin) that govern relatively slow physiological processes. Prolonged kinetics may not be as appropriate, however, for proteins that regulate processes such as glucose homeostasis, which tend to be much faster.

To address this potential limitation of the transcriptional-switch system, Rivera et al. (2000), recently developed a second technology that allows protein secretion from the endoplasmic reticulum (ER) to be rapidly regulated (Figure C.4). Therapeutic proteins are expressed as fusions with a conditional aggregation domain (CAD). CADs self-interact, and fusion proteins therefore form an aggregate in the ER that is far too large to be transported. Rivera and colleagues showed that the addition of cell-permeant ligand ("disaggregator") to transfected cells dissolves the aggregates and permits the rapid transport of therapeutic proteins from the ER via the constitutive secretory pathway.

To produce bioactive proteins, CAD moieties must be removed. Rivera et al. solved this problem by interposing a furin cleavage sequence between therapeutic protein and the CAD. In one example, Rivera et al. (2000) demonstrated that a natural version of hGH could be secreted in a controllable fashion using disaggregator technology. Thus, a single amino acid change (Phe^{36} to Met) converted monomeric FKBP12 into a CAD. Recombinant hGH was generated via a cDNA construct (Fig. C.3) consisting of a CMV promoter, signal sequence, four CAD motifs, a furin cleavage signal, and growth hormone (proinsulin was also used). Vectors were stably transfected into HT1080 cells and fluorescence microscopy was used to demonstrate ER retention of both insulin and growth hormone in the absence of disaggregator. Cells expressing fusion proteins were then treated with increasing concentrations of disaggregator for 2 hours. The authors showed that accumulated protein was released by disaggregator administration, and the rate of release was controllable over an ~20-fold dose range. In the absence of ligand, fusion proteins were found only in cell lysate samples, whereas 2 hours after addition of ligand, fusion proteins were cleaved appropriately and secreted, as determined by Western analysis. Finally, myoblast transfer was used to demonstrate feasibility of the system in animal models. To this end, engineered cells were implanted into mice made diabetic by treatment with streptozotocin. Administration of vehicle failed to normalize serum glucose concentrations. However, after intravenous administration of ligand insulin was detected in serum within 15 minutes and peaked by 2 hours. Indeed, 2 hours after administration of a 10.0-mg/kg dose of ligand, the circulating insulin concentration increased to greater than 200.0-pM and serum glucose decreased concomitantly to normal. Lower doses of ligand were less effective.

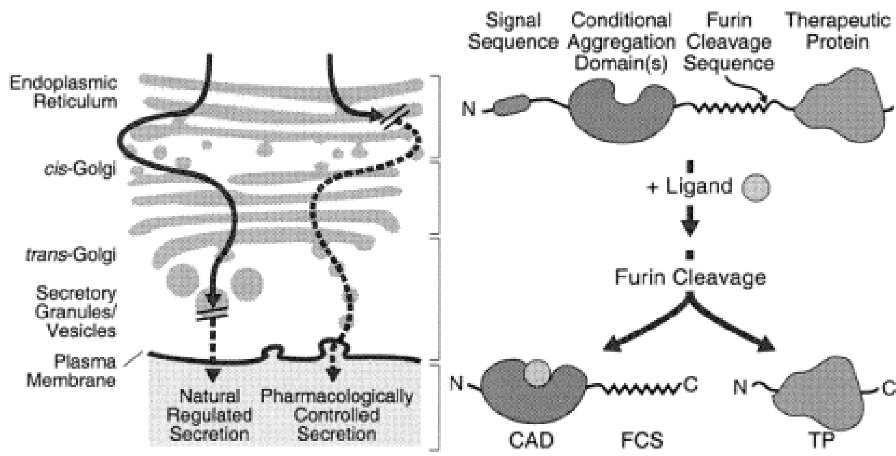


Figure C.4. Scheme for the pharmacologic control of protein secretion. (A) (*left*) Natural control of protein secretion (protein is stored in the secretory granules) is contrasted with the scheme for pharmacological control (protein is stored in the ER). (*right*) The therapeutic protein of interest (TP) is expressed as part of a fusion protein that contains, at its NH2-terminus, a signal sequence, a conditional aggregation domain (CAD), and a furin cleavage sequence (FCS). Processing and secretion of the TP is induced by ligand (Rivera et al. 2000).

Summary

Several trends have been identified:

a) Breakthroughs in controllable gene therapy technology have allowed therapeutic transgene expression to be regulated with precision over a period of months to years. The technology features low baseline transgene expression, a high induction ratio, and control via an orally available, cell permeant small molecule. Feasibility has been established in a series of elegant studies that employ recombinant adeno-associated viral (rAAV) vectors. *These breakthroughs are unique to gene therapy, i.e., similar levels of pro-drug stability and control simply do not exist for more traditional drug substances (small molecules, peptides, and proteins)*.

b) *One may see enormous improvements in patient care pathways*. For diabetes and other endocrinopathies, the standard of care may change from "multiple daily injections" to a "single injection of gene therapy followed by ingestion of multiple tablets each day." Drug therapy could truly be personalized: once individual disease patterns are established (e.g., via sensor technology), the patient and physician could work together to develop a rational, personalized regimen of small molecule administration that would be expected to yield improved compliance and better control of disease; this in turn should lessen the cost of disease to U.S. society.

c) Given the availability of a panel of cell-permeant small molecules, gene therapy becomes a combined validation/development platform in which the therapy is a stable pro-drug that remains controllable for years following initial injection of tissues such as skeletal muscle. The small molecule panel would likely form an important core element of a company's intellectual property.

d) Given the biological and technological complexity associated with genome analysis technology, an interdisciplinary spirit will be required to advance our knowledge base in basic science and drug development. Although significant technological hurdles must be traversed, the potential advantages are enormous if controllable gene therapy can realize its potential as a validation and delivery platform. Drug discovery and development may one day be routine (a more-or-less turnkey process), characterized by direct, efficient transitions *from* database query *to* rational isolation of the relevant cDNA *to* preclinical validation, *to* validation in human clinical trials (Fig. C.1). Because the "drug substance" typically will consist of a recombinant gene and a small-molecule controller, many aspects of formulation, manufacturing, biodistribution, and toxicity would be well understood *prior to* initiation of a new development program. Obviously, companies would operate in an environment of significantly reduced risk relative to the current situation; this environment would allow companies to explore a much broader range of drug targets than typically is explored today.

e) Finally, we envision a pharmaceutical industry that possesses the technological tools and economic incentives to take full advantage of the power of genomics. Specifically, the vision proposed here is that in 10 to 15 years the U.S. private sector will have a drug discovery/drug development pathway that is significantly more cost effective (more turnkey and less risky) than what we now have and is capable of taking full advantage of the promise of the human genome sequence. (Pharmaceutical companies could actually take calculated risks!) If this vision is realized, one can easily imagine how the process of technology transfer from the developed to the undeveloped world would be incentivized for the first time.

References

Auricchio, A., Hildinger, M, O'Connor, E., Gao, G.P., and Wilson, J.M. 2001. Isolation of highly infectious and pure adeno-associated virus type 2 vectors with a single-step gravity-flow column. *Hum. Gene Ther*. 12:71.

Ballas, Z.K., Rasmussen, W.L., and Krieg, A.M. 1996. Induction of NK activity in murine and human cells by CpG motifs in oligodeoxynucleotides and bacterial DNA. *J. Immunol*. 157:1840.

Bonadio, J. 2000. Tissue engineering via local gene delivery: Update and future prospects for enhancing the technology. *Advanced Drug Delivery Reviews* 44:185.

Brown, E.J., Albers, M.W., Shin, T.B., Ichikawa, K., Keith, C.T., Lane, W.S., and Schreiber, S.L. 1994. A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature* 369:756.

Davidson, B.L., Stein, C.S., Heth, J.A., Martins, I., Kotin, R.M., Derksen, T.A., Zabner, J., Ghodsi, A., Chiorini J.A. 2000. Recombinant adeno-associated virus type 2,4, and 5 vectors: transduction of variant cell types and regions in the mammalian central nervous system. *Proc. Natl. Acad. Sci*. U.S.A. 97:3428.

Deng, G-M., Nilsson, I-M., Verdrengh, M., Collins, L.V., and Tarkowski, A. 1999. Intra-articularly localized bacterial DNA containing CpG motifs induces arthritis. *Nature Med*. 5:702.

Dry, S., McCarthy, S., and Harris, T. 2001. Structural genomics in the biotechnology sector. *Nat. Biotechnol*. 29:946.

Evans, W.E., and Relling, M.V. 1999. Pharmacogenomics: translation functional genomics into rational therapeutics. *Science* 286:487.

Felgner, P.L., Barenholz, Y., Behr, J.P., Cheng, S.H., Cullis, P., Huang, L., Jessee, J.A., Seymour, L., Szoka, F., Thierry, A.R., Wagner, E., and Wu, G. 1997. Nomenclature for synthetic gene delivery systems. *Hum. Gene Ther.* 8:511.

Fields, S. 2001. Proteomics in genomeland. *Science* 291:1221.

Hernandez, Y.J., et al. 1999. Latent adeno-associated virus infection elicits humoral but not cell-mediated immune responses in a nonhuman primate model. *J. Virol.* 73:8549.

Herzog, R.W., Hagstrom, J.N., Kung, S-H., Tai, S.J., Wilson, J.M., Fisher, K.J., and High, K.A. 1997. Stable gene transfer and expression of human blood coagulation factor IX after intramuscular injection of recombinant adeno-associated virus. *Proc. Natl. Acad. Sci*. U.S.A. 94:5804.

Herzog, R.W., Yang, E.Y., Couto, L.B., Hagstron, J.N., Elwell, D., Fields, P.A., Burton, M., Bellinger, D.A., Read, M.S., Brinkhous, K.M., Podsakoff, G.M., Nichols, T.C., Kurtzman, G.J., and High, K.A. 1999. Long-term correction of canine hemophilia B by gene transfer of blood coagulation factor IX mediated by adeno-associated viral vector. *Nature Med*. 5:56.

Hol, W.G.J. 2000. Structural genomics for science and society. *Nat. Structural Biol*. 7 Suppl:964.

Ideker, T., Galitski, T., and Hood, L. 2001. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* 2:343.

Jooss, K., Yang, Y., Fisher, K.J., and Wilson, J.M. 1998. Transduction of dendritic cells by DNA viral vectors directs the immune response to transgene products in muscle fibers. *J. Virol*. 72:4212.

Kay, M.A., Glorioso, J.C., and Naldini, L. 2001. Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nature Med*. 7:33.

Kay, M.A., Manno, C.S., Ragni, M.V., Larson, P.J., Couto, L.B., McClelland, A., Glader, B., Chew, A.J., Tai, S.J., Herzog, R.W., Arruda, V., Johnson, F., Scallan, C., Skarsgard, E., Flake, A.W., and High, K.A. 2000. Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector. *Nature Genet*. 24:257.

Klinman, D.M., Yi, A.K., Beucage, S.L., Conover, J., and Krieg, A.M. 1996. CpG motifs present in bacterial DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon gamma. *Proc. Natl. Acad. Sci*. U.S.A. 93:2879.

Kozarsky, K.F., and Wilson, J.M. 1993. Gene therapy: adenovirus vectors. *Curr. Opin. Genet. Dev*. 3:499.

MacBeath, G., and Schreiber, S.L. 2000. Printing proteins as microarrays for high-throughput function determination. *Science* 289:1673.

MacColl, G., Bunn, C., Goldspink, G., Bouloux, P., and Gorecki, D.C. 2001. Intramuscular plasmid DNA injection can accelerate autoimmune responses. *Gene Ther*. 8:1354.

Magari, S.R., Rivera, V.M., Iuliucci, J.D., Gilman, M., and Cerasoli, F. 1997. Pharmacologic control of a humanized gene therapy system implanted into nude mice. *J. Clin. Invest*. 100:2865.

Mahalati, K., and Kahan, B.D. 2001. Clinical pharmacokinetics of sirolimus. *Clin. Pharmacokinet*. 40:573.

Malik, A.K., Monahan, P.E., Allen, D.L., Chen, B.G., Samulski, R.J., and Kurachi, K. 2000. Kinetics of recombinant adeno-associated virus-mediated gene transfer. J. Virol. 74:3555.

McKeown, T. 1988. *The origins of human disease*. Oxford, UK: Basil Blackwell Ltd., pp. 1-233.

Meyer, U.A., and Zanger, U.M. 1997. Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annu. Rev. Pharmacol. Toxicol*. 37:269.

Miller, D.G., Adam, M.A., and Miller, A.D. 1990. Gene Transfer by retrovirus vectors occurs only in cells that are actively replicating at the time of infection. *Mol. Cell Biol*. 10:4239.

Monahan, P.E., and Samulski, R.J. 2000. AAV vectors: is clinical success on the horizon? *Gene Ther*. 7:24.

Nateson, S., Molinari, E., Rivera, V.M., Rickles, R.J., and Gilman, M. 1999. A general strategy to enhance the potency of chimeric transcriptional activators. *Proc. Natl. Acad. Sci*. U.S.A. 96:13898.

Payette, P.J., Weeratna, R.D., McCluskie, M.J., and Davis, H.L. 2001. Immune-mediated destruction of transfected myocytes following DNA vaccination occurs via multiple mechanisms. *Gene Ther*. 8:1395.

Pollock, R., Issner, R., Zoller, K., Natesan, S., Rivera, V.M., and Clackson, T. 2000. Delivery of a stringent dimerizer-regulated gene expression system in a single retroviral vector. *Proc. Natl. Acad. Sci*. USA 97: 13221.

Rando, T.A., and Blau, H.M. 1994. Primary mouse myoblast purification, characterization, and transplantation for cell-mediated gene therapy. *J. Cell Biol*. 125:1275.

Rivera, V.M., Clackson, T., Natesan, S., Pollock, R., Amara, J.F., Keenan, T., Magari, S.R., Phillips, T., Courage, N.L, Cerasoli, F. Jr, Holt, D.A., and Gilman, M. 1996. A humanized system for pharmacologic control of gene expression. *Nat. Med*. 2:1028.

Rivera, V.M., Wang, X., Wardwell, S., Courage, N.L., Volchuk, A., Keenan, T., Holt, D.A., Gilman, M., Orci, L., Cerasoli, F. Jr., Rothman, J.E., and Clackson, T. 2000. Regulation of protein secretion through controlled aggregation in the endoplasmic reticulum. *Science* 287:826.

Rivera, V.M., Ye, X., Courage, N.L., Sachar, J., Cerasoli, F. Jr, Wilson, J.M., and Gilman, M. 1999. Long-term regulated expression of growth hormone in mice after intramuscular gene transfer. *Proc. Natl. Acad. Sci*. U.S.A. 96:8657.

Roe, T., Reynolds, T.C., Yu, G., and Brown, P.O. 1993. Integration of murine leukemia virus DNA depends on mitosis. *EMBO J*. 12:2099.

Ross-Macdonald, P., Coelho, P.S., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K.H., Sheehan, A., Symoniatis, D., Umansky, L., Heidtman, M., Nelson, F.K., Iwasaki, H., Hager, K., Gerstein, M., Miller, P., Roeder, G.S., and Snyder, M. 1999. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 402:362.

Russell, R.B., and Eggleston, D.S. 2000. New roles for structure in biology and drug discovery. *Nat. Structural Biol*. 7 Suppl:928.

Somia, N., and Verma, I.M. 2000. Gene therapy: trials and tribulations. *Nature Rev. Genetics* 1:91.

Sporn, M.B. 1999. Microbes Infect. TGF-beta: 20 years and counting. 1:1251.

Standaert, R.F., Galat, A., Verdine, G.L., and Schreiber, S.L. 1990. Molecular cloning and overexpression of the human FK506-binding protein FKBP. *Nature* 346:671.

Strausberg, R.L., and Riggins, G.J. Navigating the human transcriptome. Proc. Natl. Acad. Sci. U.S.A. 98:11837.

Velculescu, V.E., Vogelstein, B., and Kinzler, K.W. 2000. Analyzing uncharted transcriptomes with SAGE. *Trends Genet*. 16:423.

Wilson, J.F., Weale, M.E., Smith, A.C., Gratrix, F., Fletcher, B., Thomas, M.G., Bradman, N., and Goldstein, D.B. 2001. Population genetic structure of variable drug response. *Nat. Biotechnol*. 29:265.

Xiao, X., Li, J., and Samulski, R.J. 1996. Efficient long-term gene transfer into muscle tissue of immunocompetent mice by adeno-associated virus vector. *J. Virol*. 70:8098.

Yang, Y., Li, Q., Ertl, H.C., and Wilson, J.M. 1995. Cellular and humoral immune responses to viral antigens create barriers to lung-directed gene therapy with recombinant adenoviruses. *J. Virol*. 69:2004.

Ye, X., Rivera, V.M., Zoltick, P., Cerasoli, F. Jr., Schnell, M.A., Gao, G., Hughes, J.V., Gilman, M., and Wilson J.M. 1999. Regulated delivery of therapeutic proteins after in vivo somatic cell gene transfer. *Science* 283:88.

IMPLICATIONS OF THE CONTINUUM OF BIOINFORMATICS

Peter C. Johnson, TissueInformatics, Inc.

The once impenetrable complexity of biology has come face to face with rapidly expanding microprocessing power and information management solutions, and this confluence is changing our world. The parallel development of tools needed to extract biological meaning from DNA, proteins, cells, tissues, organisms, and society as a whole has set the stage for improved understanding of biological mechanisms. This is being augmented by our ability to manage this information in uniform ways and to ask questions about relationships across broad levels of biological scale. This multiscalar description of biology from the molecular to the societal, with all of the tools needed to draw correlations across its landscape, is known as the continuum of bioinformatics (COB).

Though presently immature, the COB is growing in richness daily. Driven initially by the need to manage DNA and protein sequence data, it has grown with the inclusion of cellular imaging, tissue analysis, radiological imaging, and societal healthcare informatics inputs. It is presently virtual but, like the Internet before it, it is being tied together through the development of standard systems, query tools, and security measures. As it develops, the COB is changing our world through the enhancement of our understanding of biological process and the acceleration of development of products that can benefit man, animals, and plants. The unusual precision with which biological data is represented within the COB is making it possible to reduce the degrees of freedom normally accorded biological understanding — and therefore to enable the individualization of solutions that will protect life.

Nanotechnology will play a major role in the development of information gathering and processing systems for the COB.

Definition of Bioinformatics

The science of bioinformatics presents the rich complexity of biology in such a way that meaning can be extracted using digital tools. As a discipline having multiple parts, it can be defined overall in a number of ways. One definition of bioinformatics and its components is as follows (D'Trends n.d.):

(1) *Bioinformatics* - database-like activities involving persistent sets of data that are maintained in a consistent state over essentially indefinite periods of time

(2) *Computational biology* - the use of algorithmic tools to facilitate biological analyses

(3) *Bioinformation infrastructure* - the entire collective of information management systems, analysis tools and communication networks supporting biology

This composite definition points out the importance of three activities critical to the success of bioinformatics activities:

- The use of analytic methods to enable the presentation of biological information in digital fashion.
- The leveraging of massive digital storage systems and database technologies to manage the information obtained.
- The application of digital analytic tools to identify patterns in the data that clarify causes and effects in biological systems, augmented by visualization tools that enable the human mind to rapidly grasp these patterns.

Bioinformatics makes the complexity of biological systems tangible. Taken in the stepwise fashion described above, this complexity can often be reduced to terms that are understandable to scientists probing biological problems. Biological complexity is worthwhile to understand. A clear appreciation of cause and effect in biological systems can provide the knowledge needed to develop drugs and other medical therapies and also to provide a greater appreciation for what we are as humans beings. It is interesting to note that biological complexity is so extreme that it challenges the best that high-performance computing presently has to offer. Ironically, the fulfillment of the Socratic adage "Know Thyself" can now only be achieved through man's interaction with and dependence upon computing systems.

The recent accomplishment of sequencing the human genome (and now the genomes of several other species) focused attention on the information processing requirements at the molecular end of the biological spectrum. For a time, it seemed that "bioinformatics" was wholly concerned with the management and deciphering of genetic information. Soon, information descriptive of the patterns of expression of proteins and their interactions was added (proteomics). Since this information required stratification by disease type, cellular and tissue information became important to consider. Inevitably, it became apparent that information descriptive of the whole organism, such as radiological data, other morphometric data, chemistries, and other health record data should be included. Once this was done, aggregated societal data was the next logical addition.

The picture that has come into view is therefore one of a continuum of bioinformatics (Figure C.5). In the COB model, linked data at multiple scales of biological complexity are considered together for both individuals and aggregates of individuals. The key to the value of the COB will be the ability to derive correlations between *causes* (such as gene expression, protein interactions, and the like) and *effects* (such as healthcare outcomes for individuals and societies). In this model, it may well be possible one day to determine the cost to society of the mutation of a single gene in a single individual! It will also be possible to predict with clarity which drugs will work for which individuals and why. By taking a reverse course through the COB from effects to causes, it will also be possible to sharply identify proteins that can serve as drug targets for specific disease states.

Information Capture

In order to benefit from the COB, information descriptive of biology at multiple scales must first be captured accurately and then managed such that different types of data can be interpreted with reference to one another. It is in this area that the convergence of nanotechnology and biotechnology will occur, since nanotechnology provides enabling mechanisms for the capture and management of complex biological information, particularly at the level of molecular expression data.

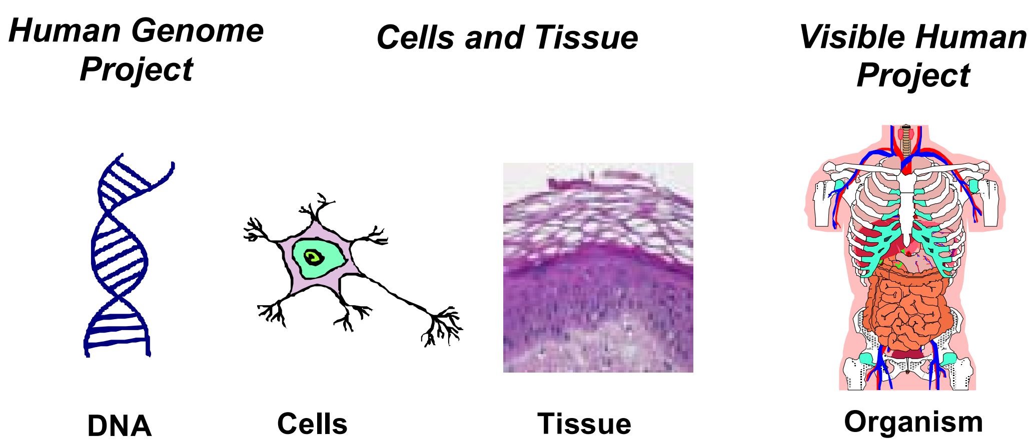


Figure C.5. Multiple scales of biological activity and form comprise the entire organism. The Continuum of Bioinformatics is an information system that includes and can correlate information from all of these scales of data. Though not shown in the figure, the aggregation of individual data into societal data (as in the form of healthcare statistics) is extremely valuable, since it places the individual's data within the context of the society as a whole.

A simple way to look at this issue is to first note that to be useful in the COB context, all biological data must first be captured using techniques that enable its ultimate conversion to digital form. The mechanisms differ, depending upon the point in the COB under consideration. Table C.3 shows the levels of the COB and the tools needed to capture data digitally at the proper level of discretion to enable computerized correlation between data pools.

Table C.3
Tools required to capture human biological information at different levels of scale, constituting the Continuum Of Bioinformatics

Biological Scale	Tools For Information Capture
DNA, Genes	DNA Sequencers Electrophoresis Affinity Microarrays
Proteins	Electrophoresis Mass Spectrometry Affinity Microarrays
Cells	Bioassays Fluorescent probes Digital Imaging
Tissues	Digital Imaging Hyperquantitative Analysis
Organism	Digital Radiology (X-Ray, CT, MRI, PET) Chemistry Data Healthcare Record
Society	Aggregated Healthcare Records

Ideally, information at all levels of scale would be captured from the same individual and then aggregated into a societal record. Since this is unpractical, aggregated information will most likely be used, and this will grow richer over time. Privacy concerns are often raised when highly discrete and potentially predictive personal information is gathered in this way. However, it is most likely that COB data (as this virtual network begins to merge together) will be anonymized sufficiently so that individuals will be protected. Indeed, one way to look at the COB is to envision it as a powerful reference database against which an individual's data can be compared in order to provide an individual with contextual information regarding his or her health at any point in time. This is the essence of what is known as "Systems Biology," as well.

Tissue Information as a Specific Instance

A specific example of the conversion of biological information to digital information occurs at the tissue level. Until recently, it was felt that only a pathologist could interpret the meaning of patterns of cells and other structural components of tissue. This meaning was summed up in the diagnosis that was applied to the tissue and used to guide healthcare decision-making. Over the past two decades, digital imaging of tissues on slides has created the basis for management of tissue information at the image level for ease of data sharing between pathologists and researchers. However, this did not convert the data completely into digital form, because human interpretation and diagnostic assignment of the overall image were still required. This limited the ability to correlate tissue data with other biological information to the level of resolution that diagnosis provided.

Recently, it has become possible to use automated machine vision analysis systems to measure all of the components that can be made visible within a tissue (both structural and functional) with reference to one another. This is known as *Hyperquantitative Analysis of Tissue* (Fig. C.6).

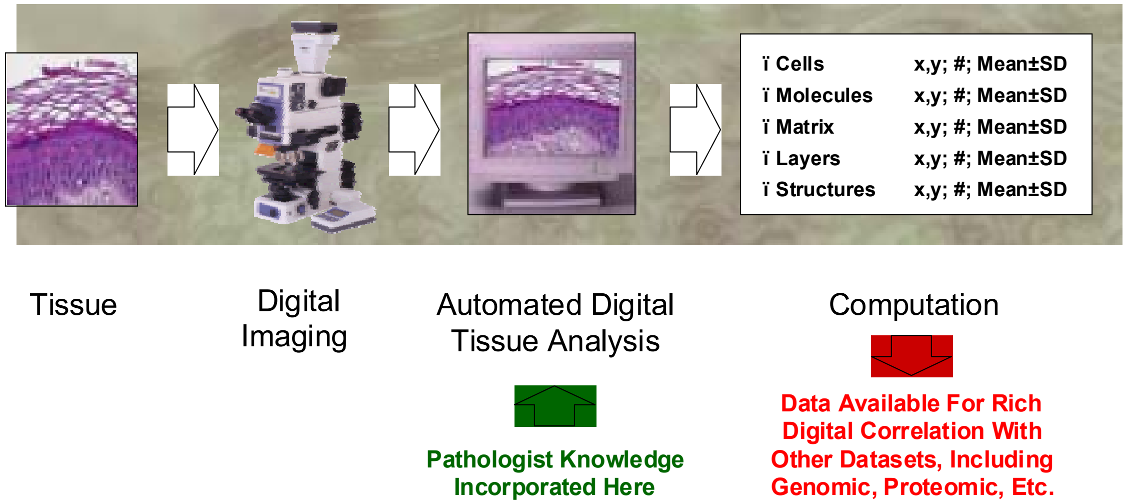


Figure C.6. Capture of tissue information in hyperquantitative fashion. All components of the tissue that can be made visible are located simultaneously after robotic capture of slide-based images. This step automates the analysis of tissue, putting it immediately into a form that enables sharing of images and derived data.

Preparation of tissue information in this way requires two steps:

a. automated imaging that enables location of tissue on a microscope slide and the capture of a composite image of the entire tissue — or tissues — on the slide

b. the application of image analytic software that has been designed to automatically segregate and co-localize in Cartesian space the visible components of tissue (including molecular probes, if applied)

Tissue information captured in this way enables very precise *mathematical comparison* of tissues to detect change (as in toxicology testing or, ultimately, clinical diagnostics). In each case, substantial work must first be done to collect normative reference data from tissue populations of interest.

More importantly, when tissue information is reduced to this level of scale, the data is made available for more precise *correlation* with other data sets in the continuum of bioinformatics in the following applications:

- *Backward correlation:* "Sorter" of genomic and proteomic data

  Rationale*:* When gene or protein expression data are culled from a tissue that has undergone hyperquantitative analysis, tighter correlations are possible between molecular expression patterns and tissue features whose known biological roles help to explain the mechanisms of disease — and therefore may help to identify drug targets more sharply.

- *Forward correlation:* Stratifier of diagnosis with respect to prognosis

  Rationale*:* When tissue information is collected along with highly detailed clinical descriptions and outcome data, subtle changes in tissue feature patterns within a diagnostic group may help to further stratify prognoses associated with a diagnosis and may prompt more refined diagnostic classifications.

- *Pan Correlation:* Tighten linkage of prognosis with molecular diagnostics

Rationale*:* Since tissue is the classical "site of diagnosis," the use of tissue information to correlate with molecular expression data and clinical outcome data validates those molecular expression patterns with reference to their associated diseases, enabling their confident application as molecular diagnostics.

Nanotechnology developments applicable to imaging and computational science will aid and abet these discoveries.

Information Management

The physical management of the large volumes of information needed to represent the COB is essentially an information storage and retrieval problem. Although only several years ago the amount of information that required management would have been a daunting problem, this is far less so today. Extremely large storage capacities in secure and fast computer systems are now commercially available. While excellent database systems are also available, none has yet been developed that completely meets the needs of the COB as envisioned. Database system development will continue to be required in order for the COB to be applied maximally. Several centers are now attempting the development of representative databases of this type.

Extracting Value From the Continuum of Bioinformatics

Once the COB is constructed and its anonymized data becomes available, it can be utilized by academia, industry, and government for multiple critical purposes. Table C.4 shows a short list of applications.

Table C.4
Applications of the COB in multiple sectors

Academic Applications	• Education • Research
Industrial Applications	• Drug Development • Medical Device Development • Tissue Engineering • Marketing
Government Applications	• Population Epidemiology • Disease Tracking • Healthcare Cost Management

In order for COB data to be put to best use, considerable work will be needed to incorporate statistical methodology and robust graphical user interfaces into the COB. In some cases, the information gleaned will be so complex that new methods of visualization of data will need to be incorporated. The human mind is a powerful interpreter of graphical patterns. This may be the reason why tissue data — classically having its patterns interpreted visually by a pathologist — was the last in the continuum to be reduced to discrete digital form.

As the COB develops, we are likely to see novel data visualization methods applied in ways that cannot be envisioned at all today. In each instance, the robustness of these tools will ultimately depend on the validity of the data that was entered into the COB and on the mode of application of statistical tools to the data being analyzed.

Impact on Human Health

The COB will significantly enhance our ability to put individual patterns of health and disease in context with that of the entire population. It will also enable us to better understand the mechanisms of disease, how disease extends throughout the population, and how it may be better treated. The availability of the COB will resect time and randomness from the process of scientific hypothesis testing, since data will be available in a preformed state to answer a limitless number of questions. Finally, the COB will enable the prediction of healthcare costs more accurately. All of these beneficial reesults will be accelerated through the application of nanotechnology principles and techniques to the creation and refinement of imaging, computational, and sensing technologies.

Reference

D'Trends, Inc. http://www.d-trends.com/Bioinformatics/bioinformatics.html.

West, J.L., and N.J. Halas. 2000. Applications of nanotechnology to biotechnology commentary, *Curr. Opin. Biotechnol.* 11(2):215-7 (Apr.).

SENSORY REPLACEMENT AND SENSORY SUBSTITUTION: OVERVIEW AND PROSPECTS FOR THE FUTURE

Jack M. Loomis, University of California, Santa Barbara

The traditional way of dealing with blindness and deafness has been some form of sensory substitution — allowing a remaining sense to take over the functions lost as the result of the sensory impairment. With visual loss, hearing and touch naturally take over as much as they can, vision and touch do the same for hearing, and in the rare cases where both vision and hearing are absent (e.g., Keller 1908), touch provides the primary contact with the external world. However, because unaided sensory substitution is only partially effective, humans have long improvised with artifices to facilitate the substitution of one sense with another. For blind people, braille has served in the place of visible print, and the long cane has supplemented spatial hearing in the sensing of obstacles and local features of the environment. For deaf people, lip reading and sign language have substituted for the loss of speech reception. Finally, for people who are both deaf and blind, fingerspelling by the sender in the palm of the receiver (Jaffe 1994; Reed et al. 1990) and the Tadoma method of speech reception (involving placement of the receiver's hand over the speaker's face) have provided a means by which they can receive messages from others (Reed et al. 1992).

Assistive Technology and Sensory Substitution

Over the last several decades, a number of new assistive technologies, many based on electronics and computers, have been adopted as more effective ways of promoting sensory substitution. This is especially true for ameliorating blindness. For example, access to print and other forms of text has been improved with these technologies: electronic braille displays, vibtrotactile display of optically sensed print (Bliss et al. 1970), and speech display of text sensed by video camera (Kurzweil 1989). For obstacle avoidance and sensing of the local environment, a number of ultrasonic sensors have been developed that use either auditory or tactile displays (Brabyn 1985; Collins 1985; Kay 1985). For help with large-scale wayfinding, assistive technologies now include electronic signage, like the system of Talking Signs (Crandall et al. 1993; Loughborough 1979; see also http://www.talkingsigns.com/), and navigation systems relying on the Global Positioning System (Loomis et al. 2001), both of which make use of auditory displays. For deaf people, improved access to spoken language has been made possible by automatic speech recognition coupled with visible display of text; in addition, research has been conducted on vibrotactile speech displays (Weisenberger et al. 1989) and synthetic visual displays of sign language (Pavel et al. 1987). Finally, for deaf-blind people, exploratory research has been conducted with electromechanical Tadoma displays (Tan et al. 1989) and finger spelling displays (Jaffe 1994).

Interdisciplinary Nature of Research on Sensory Replacement / Sensory Substitution

This paper is concerned with compensating for the loss of vision and hearing by way of sensory replacement and sensory substitution, with a primary focus on the latter. Figure C.7 shows the stages of processing from stimulus to perception for vision, hearing, and touch (which often plays a role in substitution) and indicates the associated basic sciences involved in understanding these stages of processing. (The sense of touch, or haptic sense, actually comprises two submodalities: kinesthesis and the cutaneous sense [Loomis and Lederman 1986]; here we focus on mechanical stimulation). What is clear is the extremely interdisciplinary nature of research to understand the human senses. Not surprisingly, the various attempts to use high technology to remedy visual and auditory impairments over the years have reflected the current scientific understanding of these senses at the time. Thus, there has been a general progression of technological solutions starting at the distal stages (front ends) of the two modalities, which were initially better understood, to solutions demanding an understanding of the brain and its functional characteristics, as provided by neuroscience and cognitive science.

	Scientific discipline(s)	**Vision**	**Hearing**	**Touch**
Cognitive processing	Cognitive Science/ Neuroscience	Multiple brain areas	Multiple brain areas	Multiple brain areas
Sensory processing	Psychophysics/ Neuroscience	Visual pathway	Auditory pathway	Somatosensory pathway
Transduction	Biophysics/Biology	Retina	Cochlea	Mechanoreception
Conduction	Physics/Biology	Optics of eye	Outer/middle ears	Skin
Stimulus	Physics	Light	Sound	Force

Figure C.7. Sensory modalities and related disciplines.

Sensory Correction and Replacement

In certain cases of sensory loss, sensory correction and replacement are alternatives to sensory substitution. Sensory correction is a way to remedy sensory loss prior to transduction, the stage at which light or sound is converted into neural activity (Figure C.7). Optical correction, such as eyeglasses and contact lenses, and surgical correction, such as radial keratotomy (RK) and laser in situ keratomileusis (LASIK), have been employed over the years to correct for refractive errors in the optical media prior to the retina. For more serious deformations of the optical media, surgery has been used to restore vision (Valvo 1971). Likewise, hearing aids have long been used to correct for conductive inefficiencies prior to the cochlea. Because our interest is in more serious forms of sensory loss that cannot be overcome with such corrective measures, the remainder of this section will focus on sensory replacement using bionic devices.

In the case of deafness, tremendous progress has already been made with the cochlear implant, which involves replacing much of the function of the cochlea with direct electrical stimulation of the auditory nerve (Niparko 2000; Waltzman and Cohen 2000). In the case of blindness, there are two primary approaches to remedying blindness due to sensorineural loss: retinal and cortical prostheses. A retinal prosthesis involves electrically stimulating retinal neurons beyond the receptor layer with signals from a video camera (e.g., Humayun and de Juan 1998); it is feasible when the visual pathway beyond the receptors is intact. A cortical prosthesis involves direct stimulation of visual cortex with input driven by a video camera (e.g., Normann 1995). Both types of prosthesis present enormous technical challenges in terms of implanting the stimulator array, power delivery, avoidance of infection, and maintaining long-term effectiveness of the stimulator array.

There are two primary advantages of retinal implants over cortical implants. The first is that in retinal implants, the sensor array will move about within the mobile eye, thus maintaining the normal relationship between visual sensing and eye movements, as regulated by the eye muscle control system. The second is that in retinal implants, connectivity with the multiple projection centers of the brain, like primary visual cortex and superior colliculus, is maintained without the need for implants at multiple sites. Cortical implants, on the other hand, are technically more feasible (like the delivery of electrical power), and are the only form of treatment for blindness due to functional losses distal to visual cortex. For a discussion of other pros and cons of retinal and cortical prostheses, visit the Web site (http://insight.med.utah.edu/research/normann/normann.htm) of Professor Richard Normann of the University of Utah.

Interplay of Science and Technology

Besides benefiting the lives of blind and deaf people, information technology in the service of sensory replacement and sensory substitution will continue to play another very important role — contributing to our understanding of sensory and perceptual function. Because sensory replacement and sensory substitution involve modified delivery of visual and auditory information to the perceptual processes in the brain, the way in which perception is affected or unaffected by such modifications in delivery is informative about the sensory and brain processes involved in perception. For example, the success or lack thereof of using visual displays to convey the information in the acoustic speech signal provides important clues about which stages of processing are most critical to effective speech reception. Of course, the benefits flow in the opposite direction as well: as scientists learn more about the sensory and brain processes involved in perception, they can then use the knowledge gained to develop more effective forms of sensory replacement and substitution.

Sensory Replacement and the Need for Understanding Sensory Function

To the layperson, sensory replacement might seem conceptually straightforward — just take an electronic sensor (e.g., microphone or video camera) and then use its amplified signal to drive an array of neurons somewhere within the appropriate sensory pathway. This simplistic conception of "sensory organ replacement" fails to recognize the complexity of processing that takes place at the many stages of processing in the sensory pathway. Take the case of hearing. Replacing an inoperative cochlea involves a lot more than taking the amplified signal from a microphone and using it to stimulate a collection of auditory nerve fibers. The cochlea is a complex transducer that plays sound out in terms of frequency along the length of the cochlea. Thus, the electronic device that replaces the inoperative cochlea must duplicate its sensory function. In particular, the device needs to perform a running spectral analysis of the incoming acoustic signal and then use the intensity and phase in the various frequency channels to drive the appropriate auditory nerve fibers. This one example shows how designing an effective sensory replacement begs detailed knowledge about the underlying sensory processes. The same goes for cortical implants for blind people. Simply driving a large collection of neurons in primary visual cortex by signals from a video camera after a simple spatial sorting to preserve retinotopy overlooks the preprocessing of the photoreceptor signals being performed by the intervening synaptic levels in the visual pathway. The most effective cortical implant will be one that stimulates the visual cortex in ways that reflect the normal preprocessing performed up to that level, such as adaptation to the prevailing illumination level.

Sensory Substitution: An Analytic Approach

If sensory replacement seems conceptually daunting, it pales in comparison with sensory substitution. With sensory substitution, the goal is to substitute one sensory modality that is impaired or nonfunctioning with another intact modality (Bach-y-Rita 1972). It offers several advantages over sensory replacement: (1) Sensory substitution is suitable even for patients suffering sensory loss because of cortical damage and (2) because the interface with the substituting modality involves normal sensory stimulation, there are no problems associated with implanting electrodes. However, because the three spatial modalities of vision, hearing, and touch differ greatly in terms of their processing characteristics, the hope that one modality, aided by some single device, can simply assume all of the functions of another is untenable. Instead, a more reasonable expectation is that one modality can only substitute for another in performance of certain limited functions (e.g., reading of print, obstacle avoidance, speech reception). Indeed, research and development in the field of sensory substitution has largely proceeded with the idea of restoring specific functions rather than attempting to achieve wholesale substitution. A partial listing follows of the functions performed by vision and hearing, which are potential goals for sensory substitution:

- **Some functions of vision = potential goals for sensory substitution**
  - access to text (e.g., books, recipes, assembly instructions, etc.)
  - access to static graphs/pictures
  - access to dynamic graphs/pictures (e.g., animations, scientific visualization)
  - access to environmental information (e.g., business establishments and their locations)
  - obstacle avoidance
  - navigation to remote locations
  - controlling dynamic events in 3-D (e.g., driving, sports)
  - access to social signals (e.g., facial expressions, eye gaze, body gestures)
  - visual aesthetics (e.g., sunset, beauty of a face, visual art)
- **Some functions of audition = potential goals for sensory substitution**
  - access to signals and alarms (e.g., ringing phone, fire alarm)
  - access to natural sounds of the environment
  - access to denotative content of speech
  - access to expressive content of speech
  - aesthetic response to music

An analytic approach to using one sensory modality (henceforth, the "receiving modality") to take over a function normally performed by another is to (1) identify what optical, acoustic, or other information (henceforth, the "source information") is most effective in enabling that function and (2) to determine how to transform the source information into sensory signals that are effectively coupled to the receiving modality.

The first step requires research to identify what source information is necessary to perform a function or range of functions. Take, for example, the function of obstacle avoidance. A person walking through a cluttered environment is able to avoid bumping into obstacles, usually by using vision under sufficient lighting. Precisely what visual information or other form of information (e.g., ultrasonic, radar) best affords obstacle avoidance? Once one has identified the best information to use, one is then in a position to address the second step.

Sensory Substitution: Coupling the Required Information to the Receiving Modality

Coupling the source information to the receiving modality actually involves two different issues: sensory bandwidth and the specificity of higher-level representation. After research has determined the information needed to perform a task, it must be determined whether the sensory bandwidth of the receiving modality is adequate to receive this information. Consider the idea of using the tactile sense to substitute for vision in the control of locomotion, such as driving. Physiological and psychophysical research reveals that the sensory bandwidth of vision is much greater than the bandwidth of the tactile sense for any circumscribed region of the skin (Loomis and Lederman 1986). Thus, regardless of how optical information is transformed for display onto the skin, it seems unlikely that the bandwidth of tactile processing is adequate to allow touch to substitute for this particular function. In contrast, other simpler functions, such as detecting the presence of a bright flashing alarm signal, can be feasibly accomplished using tactile substitution of vision.

Even if the receiving modality has adequate sensory bandwidth to accommodate the source information, this is no guarantee that sensory substitution will be successful, because the higher-level processes of vision, hearing, and touch are highly specialized for the information that typically comes through those modalities. A nice example of this is the difficulty of using vision to substitute for hearing in deaf people. Even though vision has greater sensory bandwidth than hearing, there is yet no successful way of using vision to substitute for hearing in the reception of the raw acoustic signal (in contrast to sign language, which involves the production of visual symbols by the speaker). Evidence of this is the enormous challenge in deciphering an utterance represented by a speech spectrogram. There is the celebrated case of Victor Zue, an engineering professor who is able to translate visual speech spectrograms into their linguistic descriptions. Although his skill is an impressive accomplishment, the important point here is that enormous effort is required to learn this skill, and decoding a spectrogram of a short utterance is very time-consuming. Thus, the difficulty of visually interpreting the acoustic speech signal suggests that presenting an isomorphic representation of the acoustic speech signal does not engage the visual system in a way that facilitates speech processing.

Presumably there are specialized mechanisms in the brain for extracting the invariant aspects of the acoustic signal; these invariant aspects are probably articulatory features, which bear a closer correspondence with the intended message. Evidence for this view is the relative success of the Tadoma method of speech reception (Reed et al. 1992). Some deaf-blind individuals are able to receive spoken utterances at nearly normal speech rates by placing a hand on the speaker's face. This direct contact with articulatory features is presumably what allows the sense of touch to substitute more effectively than visual reception of an isomorphic representation of the speech signal, despite the fact that touch has less sensory bandwidth than vision (Reed et al. 1992).

Although we now understand a great deal about the sensory processing of visual, auditory, and haptic perception, we still have much to learn about the perceptual/cognitive representations of the external world created by each of these senses and the cortical mechanisms that underlie these representations. Research in cognitive science and neuroscience will produce major advances in the understanding of these topics in the near future. Even now, we can identify some important research themes that are relevant to the issue of coupling information normally sensed by the impaired modality with the processing characteristics of the receiving modality.

Achieving Sensory Substitution Through Abstract Meaning

Prior to the widespread availability of digital computers, the primary approach to sensory substitution using electronic devices was to use analog hardware to map optical or acoustic information into one or isomorphic dimensions of the receiving modality (e.g., using video to sense print or other high contrast 2-D images and then displaying isomorphic tactile images onto the skin surface). The advent of the digital computer has changed all this, for it allows a great deal of signal processing of the source information prior to its display to the receiving modality. There is no longer the requirement that the displayed information be isomorphic to the information being sensed. Taken to the extreme, the computer can use artificial intelligence algorithms to extract the "meaning" of the optical, acoustic, or other information needed for performance of the desired function and then display this meaning by way of speech or abstract symbols.

One of the great success stories in sensory substitution is the development of text-to-speech devices for the visually impaired (Kurzweil 1989). Here, printed text is converted by optical character recognition into electronic text, which is then displayed to the user as synthesized speech. In a similar vein, automatic speech recognition and the visual display of text may someday provide deaf people with immediate access to the speech of any desired interactant. One can also imagine that artificial intelligence may someday provide visually impaired people with detailed verbal descriptions of objects and their layout in the surrounding environment. However, because inculcating such intelligence into machines has proven far more challenging than was imagined several decades ago, exploiting the intelligence of human users in the interpretation of sensory information will continue to be an important approach to sensory substitution. The remaining research themes deal with this more common approach.

Amodal Representations

For 3-D space perception (e.g., perception of distance) and spatial cognition (e.g., large-scale navigation), it is quite likely that vision, hearing, and touch all feed into a common area of the brain, like the parietal cortex, with the result that the perceptual representations created by these three modalities give rise to amodal representations. Thus, seeing an object, hearing it, or feeling it with a stick, may all result in the same abstract spatial representation of its location, provided that its perceived location is the same for the three senses. Once an amodal representation has been created, it then might be used to guide action or cognition in a manner that is independent of the sensory modality that gave rise to it (Loomis et al. 2002). To the extent that two sensory modalities do result in shared amodal representations, there is immediate potential for one modality substituting for the other with respect to functions that rely on the amodal representations. Indeed, as mentioned at the outset of this chapter, natural sensory substitution (using touch to find objects when vision is impaired) exploits this very fact. Clearly, however, an amodal representation of spatial layout derived from hearing may lack the detail and precision of one derived from vision because the initial perceptual representations differ in the same way as they do in natural sensory substitution.

Intermodal Equivalence: Isomorphic Perceptual Representations

Another natural basis for sensory substitution is isomorphism of the perceptual representations created by two senses. Under a range of conditions, visual and haptic perception result in nearly isomorphic perceptual representations of 2-D and 3-D shape (Klatzky et al. 1993; Lakatos and Marks 1999; Loomis 1990; Loomis et al. 1991). The similar perceptual representations are probably the basis both for cross-modal integration, where two senses cooperate in sensing spatial features of an object (Ernst et al. 2001; Ernst and Banks 2002; Heller et al. 1999), and for the ease with which subjects can perform cross-modal matching, that is, feeling an object and then recognizing it visually (Abravanel 1971; Davidson et al. 1974). However, there are interesting differences between the visual and haptic representations of objects (e.g., Newell et al. 2001), differences that probably limit the degree of cross-modal transfer and integration. Although the literature on cross-modal integration and transfer involving vision, hearing, and touch goes back years, this is a topic that is receiving renewed attention (some key references: Ernst and Banks 2002; Driver and Spence 1999; Heller et al. 1999; Martino and Marks 2000; Massaro and Cohen 2000; Welch and Warren 1980).

Synesthesia

For a few rare individuals, synesthesia is a strong correlation between perceptual dimensions or features in one sensory modality with perceptual dimensions or features in another (Harrison and Baron-Cohen 1997; Martino and Marks 2001). For example, such an individual may imagine certain colors when hearing certain pitches, may see different letters as different colors, or may associate tactual textures with voices. Strong synesthesia in a few rare individuals cannot be the basis for sensory substitution; however, much milder forms in the larger population, indicating reliable associations between intermodal dimensions that may be the basis for cross-modal transfer (Martino and Marks 2000), might be exploited to produce more compatible mappings between the impaired and substiting modalities. For example, Meijer (1992) has developed a device that uses hearing to substitute for vision. Because the natural correspondence between pitch and elevation is space (e.g., high-pitched tones are associated with higher elevation), the device uses the pitch of a pure tone to represent the vertical dimension of a graph or picture. The horizontal dimension of a graph or picture is represented by time. Thus, a graph portraying a 45º diagonal straight line is experienced as a tone of increasing pitch as a function of time. Apparently, this device is successful for conveying simple 2-D patterns and graphs. However, it would seem that images of complex natural scenes would result in a cacophony of sound that would be difficult to interpret.

Multimodal Sensory Substitution

The discussion of sensory substitution so far has assumed that the source information needed to perform a function or functions is displayed to a single receiving modality, but clearly there may be value in using multiple receiving modalities. A nice example is the idea of using speech and audible signals together with force feedback and vibrotactile stimulation from a haptic mouse to allow visually impaired people to access information about 2-D graphs, maps, and pictures (Golledge 2002, this volume). Another aid for visually impaired people is the “Talking Signs” system of electronic signage (Crandall et al. 1993), which includes transmitters located at points of interest in the environment that transmit infrared signals carrying speech information about the points of interest. The user holds a small receiver in the hand that receives the infrared signal when pointed in the direction of the transmitter; the receiver then displays the speech utterance by means of a speaker or earphone. In order to localize the transmitter, the user rotates the receiver in the hand until receiving the maximum signal strength; thus, haptic information is used to orient toward the transmitter, and speech information conveys the identity of the point of interest.

Rote Learning Through Extensive Exposure

Even when there is neither the possibility of extracting meaning using artificial intelligence algorithms nor the possibility of mapping the source information in a natural way onto the receiving modality, effective sensory substitution is not completely ruled out. Because human beings, especially when they are young, have a large capacity for learning complex skills, there is always the possibility that they can learn mappings between two sensory modalities that differ greatly in their higher-level interpretative mechanisms (e.g., use of vision to apprehend complex auditory signals or of hearing to apprehend complex 2-D spatial images). As mentioned earlier, Meijer (1992) has developed a device (The vOICe) that converts 2-D spatial images into time-varying auditory signals. While based on the natural correspondence between pitch and height in a 2-D figure, it seems unlikely that the higher-level interpretive mechanisms of hearing are suited to handling complex 2-D spatial images usually associated with vision. Still, it is possible that if such a device were used by a blind person from very early in life, the person might develop the equivalent of rudimentary vision. On the other hand, the previously discussed example of the difficulty of visually interpreting speech spectrograms is a good reason not to base one's hope too much on this capacity for learning.

Brain Mechanisms Underlying Sensory Substitution and Cross-Modal Transfer

In connection with his seminal work with the Tactile Vision Substitution System, which used a video camera to drive an electrotactile display, Bach-y-Rita (1967, 1972) speculated that the functional substitution of vision by touch actually involved a reorganization of the brain, whereby the incoming somatosensory input came to be linked to and analyzed by visual cortical areas. Though a radical idea at the time, it has recently received confirmation by a variety of studies involving brain imaging and transcranial magnetic stimulation (TMS). For example, research has shown that (1) the visual cortex of skilled blind readers of braille is activated when they are reading braille (Sadata et al. 1996), (2) TMS delivered to the visual cortex can interfere with the perception of braille in similar subjects (Cohen et al. 1997), and (3) that the visual signals of American Sign Language activate the speech areas of deaf subjects (Neville et al. 1998).

Future Prospects for Sensory Replacement and Sensory Substitution

With the enormous increases in computing power, the miniaturization of electronic devices (nanotechnology), the improvement of techniques for interfacing electronic devices with biological tissue, and increased understanding of the sensory pathways, the prospects are great for significant advances in sensory replacement in the coming years. Similarly, there is reason for great optimism in the area of sensory substitution. As we come to understand the higher level functioning of the brain through cognitive science and neuroscience research, we will know better how to map source information into the remaining intact senses. Perhaps even more important will be breakthroughs in technology and artificial intelligence. For example, the emergence of new sensing technologies, as yet unknown, just as the Global Positioning System was unknown several decades ago, will undoubtedly provide blind and deaf people with access to new types of information about the world around them. Also, the increasing power of computers and increasing sophistication of artificial intelligence software will mean that computers will be increasingly able to use this sensed information to build representations of the environment, which in turn can be used to inform and guide visually impaired people using synthesized speech and spatial displays. Similarly, improved speech recognition and speech understanding will eventually provide deaf people better communication with others who speak the same or even different languages. Ultimately, sensory replacement and sensory substitution may permit people with sensory impairments to perform many activities that are unimaginable today and to enjoy a wide range of experiences that they are currently denied.

References

Abravanel, E. 1971. Active detection of solid-shape information by touch and vision. *Perception & Psychophysics*, 10, 358-360.

Bach-y-Rita, P. 1967. Sensory plasticity: Applications to a vision substitution system. *Acta Neurologica Scandanavica*, 43, 417-426.

Bach-y-Rita, P. 1972. *Brain mechanisms in sensory substitution*. New York: Academic Press.

Bliss, J.C., M.H. Katcher, C.H. Rogers, and R.P. Shepard. 1970. Optical-to-tactile image conversion for the blind. *IEEE Transactions on Man-Machine Systems*, MMS-11, 58-65.

Brabyn, J.A. 1985. A review of mobility aids and means of assessment. In *Electronic spatial sensing for the blind*, D.H. Warren and E.R. Strelow, eds. Boston: Martinus Nijhoff.

Cohen, L.G., P. Celnik, A. Pascual-Leone, B. Corwell, L. Faiz, J. Dambrosia, M. Honda, N. Sadato, C. Gerloff, M.D. Catala, and M. Hallett, M. 1997. Functional relevance of cross-modal plasticity in blind humans. *Nature,* 389: 180-183.

Collins, C.C. 1985. On mobility aids for the blind. In *Electronic spatial sensing for the blind*, D.H. Warren and E.R. Strelow, eds. Boston: Martinus Nijhoff.

Crandall, W., W. Gerrey, and A. Alden. 1993. Remote signage and its implications to print-handicapped travelers. *Proceedings: Rehabilitation Engineering Society of North America RESNA Annual Conference*, Las Vegas, June 12-17, 1993, pp. 251-253.

Davidson, P.W., S. Abbott, and J. Gershenfeld. 1974. Influence of exploration time on haptic and visual matching of complex shape. *Perception and Psychophysics*, 15 : 539-543.

Driver, J., and C. Spence. 1999. Cross-modal links in spatial attention. In *Attention, space, and action: Studies in cognitive neuroscience,* G.W. Humphreys and J. Duncan, eds. New York: Oxford University Press.

Ernst, M.O. and M.S. Banks. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415: 429 - 433.

Ernst, M.O., M.S. Banks, and H.H. Buelthoff. 2000. Touch can change visual slant perception. *Nature Neuroscience* 3: 69-73.

Golledge, R.G. 2002. Spatial cognition and converging technologies. This volume.

Harrison, J., and S. Baron-Cohen. 1997. Synaesthesia: An introduction. In *Synaesthesia: Classic and contemporary readings*, S. Baron-Cohen and J.E. Harrison eds. Malden, MA: Blackwell Publishers.

Heller, M.A., J.A. Calcaterra, S.L. Green, and L. Brown. 1999. Intersensory conflict between vision and touch: The response modality dominates when precise, attention-riveting judgments are required. *Perception and Psychophysics* 61: 1384-1398.

Humayun, M.S., and E.T. de Juan, Jr. 1998. Artificial vision. *Eye* 12: 605-607.

Jaffe, D.L. 1994. Evolution of mechanical fingerspelling hands for people who are deaf-blind. *Journal of Rehabilitation Research and Development* 3: 236-244.

Kay, L. 1985. Sensory aids to spatial perception for blind persons: Their design and evaluation. In *Electronic spatial sensing for the blind*, D.H. Warren and E.R. Strelow, eds. Boston: Martinus Nijhoff.

Keller, H. 1908. *The world I live in*. New York: The Century Co.

Klatzky, R.L., J.M. Loomis, S.J. Lederman, H. Wake, and N. Fujita. 1993. Haptic perception of objects and their depictions. *Perception and Psychophysics* 54 : 170-178.

Kurzweil, R. 1989. Beyond pattern recognition. *Byte* 14: 277.

Lakatos, S., and L.E. Marks. 1999. Haptic form perception: Relative salience of local and global features. *Perception and Psychophysics* 61: 895-908.

Loomis, J.M. 1990. A model of character recognition and legibility. *Journal of Experimental Psychology: Human Perception and Performance* 16: 106-120.

Loomis, J.M., R.G. Golledge, and R.L. Klatzky. 2001. GPS-based navigation systems for the visually impaired. In *Fundamentals of wearable computers and augmented reality,* W. Barfield and T. Caudell, eds. Mahwah, NJ: Lawrence Erlbaum Associates.

Loomis, J.M., R.L. Klatzky, and S.J. Lederman. 1991. Similarity of tactual and visual picture perception with limited field of view. *Perception* 20: 167-177.

Loomis, J.M., and S.J. Lederman. 1986. Tactual perception. In K. Boff, L. Kaufman, and J. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 31.1-31.41). New York: Wiley.

Loomis, J.M., Y. Lippa, R.L. Klatzky, and R.G. Golledge. 2002. Spatial updating of locations specified by 3-D sound and spatial language. *J. of Experimental Psychology: Learning, Memory, and Cognition* 28: 335-345.

Loughborough, W. 1979. Talking lights. *Journal of Visual Impairment and Blindness* 73: 243.

Martino, G., and L.E. Marks. 2000. Cross-modal interaction between vision and touch: The role of synesthetic correspondence. *Perception* 29: 745-754.

_____. 2001. Synesthesia: Strong and weak. *Current Directions in Psychological Science* 10: 61-65.

Massaro, D.W., and M.M. Cohen. 2000. Tests of auditory-visual integration efficiency within the framework of the fuzzy logical model of perception. *Journal of the Acoustical Society of America* 108: 784-789.

Meijer, P.B.L. 1992. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* 39: 112-121.

Neville, H.J., D. Bavelier, D. Corina, J. Rauschecker, A. Karni, A. Lalwani, A. Braun, V. Clark, P. Jezzard, and R. Turner. 1998. Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. Neuroimaging of Human Brain Function, May 29-31, 1997, Irvine, CA. *Proceedings of the National Academy of Sciences* 95: 922-929.

Newell, F.N., M.O. Ernst, B.S. Tjan, and H.H. Buelthoff. 2001. Viewpoint dependence in visual and haptic object recognition. *Psychological Science* 12: 37-42.

Niparko, J.K. 2000. *Cochlear implants: Principles and practices*. Philadelphia: Lippincott Williams & Wilkins.

Normann, R.A. 1995. Visual neuroprosthetics: Functional vision for the blind. *IEEE Engineering in Medicine and Biology Magazine* 77-83.

Pavel, M., G. Sperling, T. Riedl, and A. Vanderbeek. 1987. Limits of visual communication: The effect of signal-to-noise ratio on the intelligibility of American Sign Language. *Journal of the Optical Society of America, A* 4: 2355-2365.

Reed, C.M., L.A. Delhorne, N.I. Durlach, and S.D. Fischer. 1990. A study of the tactual and visual reception of fingerspelling. *Journal of Speech and Hearing Research* 33: 786-797.

Reed, C.M., W.M. Rabinowitz, N.I. Durlach, L.A. Delhorne, L.D. Braida, J.C. Pemberton, B.D. Mulcahey, and D.L. Washington. 1992. Analytic study of the Tadoma method: Improving performance through the use of supplementary tactual displays. *Journal of Speech and Hearing Research* 35: 450-465.

Sadato, N., A. Pascual-Leone, J. Grafman, V. Ibanez, M-P Deiber, G. Dold, and M. Hallett. 1996. Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* 380: 526-528.

Tan, H.Z., W.M. Rabinowitz, and N.I. Durlach. 1989. Analysis of a synthetic Tadoma system as a multidimensional tactile display. *Journal of the Acoustical Society of America* 86: 981-988.

Valvo, A. 1971. *Sight restoration after long-term blindness: the problems and behavior patterns of visual rehabilitation,* L L. Clark and Z.Z. Jastrzembska, eds.. New York, American Foundation for the Blind.

Waltzman, S.B., and N.L. Cohen. 2000. *Cochlear implants.* New York: Thieme.

Weisenberger, J.M., S.M. Broadstone, and F.A. Saunders. 1989. Evaluation of two multichannel tactile aids for the hearing impaired. *Journal of the Acoustical Society of America* 86: 1764-1775.

Welch, R.B., and D.H. Warren. 1980. Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin* 88: 638-667.

VISION STATEMENT: INTERACTING BRAIN

Britton Chance, University of Pennsylvania, and Kyung A. Kang, University of Louisville

Brain functional studies are currently performed by several instruments, most having limitations at this time. PET and SPECT use labeled glucose as an indicator of metabolic activity; however, they may not be used within a short time interval and also can be expensive. MRI is a versatile brain imaging technique, but is highly unlikely to be "wearable." MEG is an interesting technology to measure axon-derived currents with a high accuracy at a reasonable speed; this still requires minimal external magnetic fields, and a triply shielded micro-metal cage is required for the entire subject. While thermography has some advantages, the penetration is very small, and the presence of overlying tissues is a great problem. Many brain responses during cognitive activities may be recognized in terms of changes in blood volume and oxygen saturation at the brain part responsible. Since hemoglobin is a natural and strong optical absorber, changes in this molecule can be monitored by near infrared (NIR) detection method very effectively without applying external contrast agents (Chance, Kang, and Sevick 1993). NIR can monitor not only the blood volume changes (the variable that most of the currently used methods are measuring) but also hemoglobin saturation (the variable that provides the actual energy usage) (Chance, Kang, and Sevick 1993;Hoshe et al. 1994; Chance et al 1998). Among the several brain imagers, the "NIR Cognoscope" (Figure C.8) is one of a few that have wearability (Chance et al. 1993; Luo, nioka, and Chance 1996; Chance et al 1998). Also, with fluorescent-labeled neuroreceptors or metabolites (such as glucose), the optical method will have a similar capability for metabolic activities as PET and SPECT (Kang et al. 1998).

Nanotechnology and information technology (IT) can be invaluable for the development of future optical cognitive instruments. Nano-biomarkers targeted for cerebral function representing biomolecules will enable us to pinpoint the areas responsible for various cognitive activities as well as to diagnose various brain disorders. Nano-sized sources and detectors operated by very long lasting nano-sized batteries will be also very useful for unobstructed studies of brain function. It is important to acknowledge that in the process of taking cognitive function measurements, the instrument itself or the person who conducts the measurements should not (or should minimally) interfere with or distract the subject's cognitive activities. The ultimate optical system for cognitive studies, therefore, requires wireless instrumentation.

It is envisioned that once nanotech and IT are fully incorporated into the optical instrumentation, the sensing unit will be very lightweight, disposable Band-aid™ sensor/detector applicators or hats (or helmets) having no external connection. Stimuli triggering various cognitive activities can be given through a computer screen or visor with incorporating a virtual reality environment. Signal acquisition will be accomplished by telemetry and will be analyzed in real time. The needed feedback stimulus can also be created, depending on the nature of the analysis needed for further tests or treatments. Some of the important future applications of the kind of "cognoscope" described above are as follows:

1. Medical diagnosis of brain diseases (Chance, Kang, and Sevick 1993)

2. Identification of children with learning disabilities (Chance et al. 1993; Hoshe et al. 1994; Chance et al. 1998)

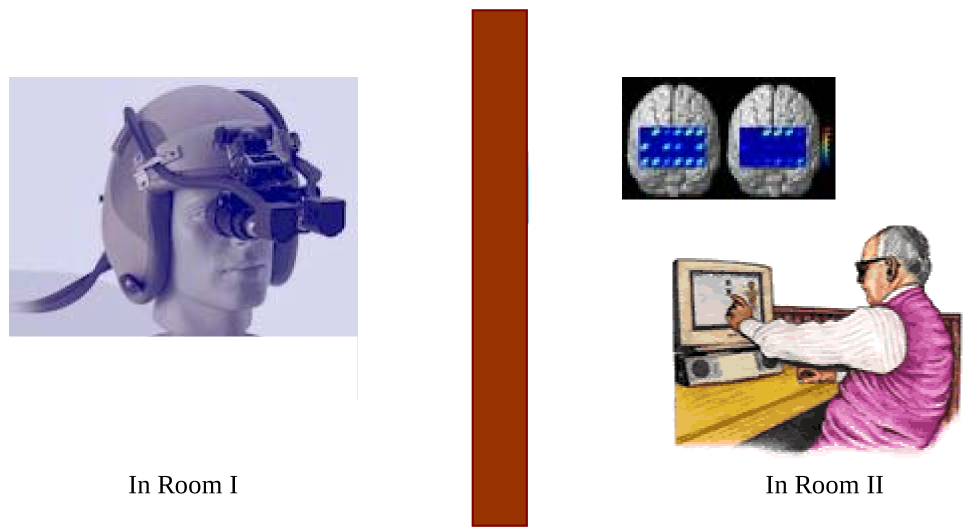
3. Assessment of effectiveness in teaching techniques (Chance et al. 1993; Hoshe et al. 1994; Heekeren et al. 1997; Chance et al. 1998)

4. Applications for cognitive science — study of the thinking process (Chance et al. 1993; Hoshe et al. 1994; Chance et al. 1998)

5. Localization of brain sites responding for various stimuli (Gratton et al. 1995; Luo, Nioka, and Chance 19997; Heekeren et al. 1997; Villringer and Chance 1997)

6. Identification of the emotional state of a human being

7. Communicating with others without going through currently used sensory systems



(a) (b)

Figure C.8. A schematic diagram of the future NIR Cognosope. (a) A wireless, hat-like multiple source-detector system can be used for brain activities while the stimulus can be given though a visor-like interactive device. While a subject can be examined (or tested) in a room (room I) without any disturbance by examiners or other non-cognitive stimuli, the examiner can obtain the cognitive response through wireless transmission, analyze the data in real-time, and also may be able to additional stimuli to the subjects for further tests, in another room (room II).

References

Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., Li, C., Overtsky, Y., Pidikiti, D., and Thomas, R., 1998. "A Novel Method for Fast Imaging of Brain Function, Noninvasively, with Light." Optical Express, 2(10): 411-423.

Chance, B., Kang, K.A., and Sevick, E., 1993. "Photon Diffusion in Breast and Brain: Spectroscopy and Imaging," *Optics and Photonics News*, 9-13.3.

Chance, B., Zhuang, Z., Chu, U., Alter, C., and Lipton, L., 1993. "Cognition Activated Low Frequency Modulation of Light Absorption in Human Brain," *PNAS*, 90: 2660-2774.

Gratton, G., Corballis, M., Cho, E., Gabiani, M., and Hood, D.C., 1995. "Shades of Gray Matter: Non-invasiveNoninvasive Optical Images of Human Brain Responses during Visual Stimulations," *Psychophysiology*, 32: 505-509.

Heekeren, H.R., Wenzel, R., Obrig, H., Ruben, J., Ndayisaba, J-P., Luo, Q., Dale, A., Nioka, S., Kohl, M., Dirnagl, U., Villringer, A., and Chance, B., 1997. "Towards Noninvasive Optical Human Brain Mapping - Improvements of the Spectral, Temporal, and Spatial Resolution of Near-infrared Spectroscopy," in Optical Tomography and Spectroscopy of Tissue: Theory, Instrumentation, Model, and Human Studies, II, Chance, B., Alfano, R., eds., *Proc. SPIE*, 2979: 847-857.

Hoshi, Y., Onoe, H., Watanabe, Y., Andersson, J., Bergstrom, M., Lilja, A., Langstom, B., and Tamura, M., 1994. "Non-synchronous Behavior of Neuronal Activity, Oxidative Metabolism and Blood Supply during Mental Tasks in Brain," *Neurosci. Lett.*, 197: 129-133.

Kang, K.A., Bruley, D.F., Londono, J.M., and Chance, B. 1998. "Localization of a Fluorescent Object in a Highly Scattering Media via Frequency Response Analysis of NIR-TRS Spectra," *Annals of Biomedical Engineering*, 26:138-145.

Luo, Q., Nioka, S., and Chance, B. 1996, "Imaging on Brain Model by a Novel Optical Probe - Fiber Hairbrush," in *Adv. Optical Imaging and Photon Migration*, Alfano, R.R., and Fumiomoto, J.G., eds., II-183-185.

Luo, Q., Nioka, S., and Chance, B. 1997. "Functional Near-infrared Image," in *Optical Tomography and Spectroscopy of Tissue: Theory, Instrumentation, Model, and Human Studies, II*, Chance, B., Alfano, R., eds., Proc. SPIE, 2979: 84-93.

Villringer, A., and Chance, B., 1997. "Noninvasive Optical Spectroscopy and Imaging of Human Brain Function," *Trends in Neuroscience*, 20: 435-442.

Focusing the Possibilities of Nanotechnology for Cognitive Evolution and Human Performance

Edgar Garcia-Rill, PhD, University of Arkansas for Medical Sciences

Two statements are advanced in this paper:

1. Nanotechnology can help drive our cognitive evolution.

2. Nanotechnology applications can help us monitor distractibility and critical judgment, allowing unprecedented improvements in human performance.

The following will provide supporting arguments for these two positions, one general and one specific, regarding applications of nanotechnology for human performance. This vision and its transforming strategy will require the convergence of nanoscience, biotechnology, advanced computing and principles in cognitive neuroscience.

Our Cognitive Evolution

How did the human brain acquire its incomparable power? Our species emerged less than 200,000 years ago, but it has no "new" modules compared to other primates. Our brains have retained vestiges from our evolutionary ancestors. The vertebrate (e.g., fish) nervous system is very old, and we have retained elements of the vertebrate brain, especially in the organization of spinal cord and brainstem systems. One radical change in evolution occurred in the transition from the aquatic to terrestrial environment. New "modules" arose to deal with the more complex needs of this environment in the form of thalamic, basal ganglia, and cortical "modules" evident in the mammalian brain. The changes in brain structure between lower and higher mammals are related to size rather than to any novel structures. There was a dramatic growth in the size of the cerebral cortex between higher mammals and monkeys. But the difference between the monkey brain, the ape brain, and the human brain is again one of size. In comparing these three brains, we find that the size of the primary cortical areas (those dealing with sensory and motor functions) are similar in size, but in higher species, secondary and especially tertiary cortical areas (those dealing with higher-level processing of sensory and motor information) are the ones undergoing dramatic increases in size, especially in the human. That is, we have conserved a number of brain structures throughout evolution, but we seem to just have more of everything, especially cortex (Donald 1991).

As individuals, the factors that determine the anatomy of our cortex are genes, environment, and enculturation (Donald 1991). For instance, the structure of the basic computational unit of the cortex, the cortical column, is set genetically. However, the connectivity between cortical columns, which brings great computational power based on experience, is set by the environment, especially during critical stages in development. Moreover, the process of enculturation determines the plastic anatomical changes that allow entire circuits to be engaged in everyday human performance. This can be demonstrated experimentally. Genetic mutations lead to dramatic deficits in function, but if there is no genetic problem yet environmental exposure is prevented (such as covering the eyes during a critical period in development) lifelong deficits (blindness) result. If both genetic and environmental factors proceed normally, but enculturation is withdrawn, symbolic skills and language fail to develop, with drastic effects.

The unprecedented growth of the cortex exposed to culture allowed us to develop more complex skills, language, and unmatched human performance. It is thought that it is our capacity to acquire symbolic skills that has led to our higher intelligence. Once we added symbols, alphabets, and mathematics, biological memory became inadequate for storing our collective knowledge. That is, the human mind became a “hybrid” structure built from vestiges of earlier biological stages, new evolutionarily-driven modules, and external (cultural “peripherals”) symbolic memory devices (books, computers, etc.), which, in turn, have altered its organization, the way we “think” (Donald 1991). That is, just as we use our brain power to continue to develop technology, that technological enculturation has an impact on the way we process information, on the way our brain is shaped. This implies that we are more complex than any creatures before, and that we may not have yet reached our final evolutionary form. Since we are still evolving, the inescapable conclusion is that nanotechnology can help drive our evolution. This should be the charge to our nanoscientists: Develop nanoscale hybrid technology.

What kind of hybird structures should we develop? It is tempting to focus nanotechnology research on brain-machine integration, to develop *implantable* devices (rather than *peripheral* devices) to “optimize” detection, perception, and responsiveness, or to increase “computational power” or memory storage. If we can ever hope to do this, we need to know how the brain processes information. Recent progress in information processing in the brain sciences, in a sense, parallels that of advances in computation. According to Moore’s Law, advances in hardware development enables a doubling of computing and storage power every 18 months, but this has not lead to similar advances in software development, as faster computers seem to encourage less efficient software (Pollack 2002, this volume). Similarly, brain research has given us a wealth of information on the hardware of the brain, its anatomical connectivity and synaptic interactions, but this explosion of information has revealed little about the software the brain uses to process information and direct voluntary movement. Moreover, there is reason to believe that we tailor our software, developing more efficient “lines of code” as we grow and interact with the environment and culture. In neurobiological terms, the architecture of the brain is determined genetically, the connectivity pattern is set by experience, and we undergo plastic changes throughout our lives in the process of enculturation. Therefore, we need to hone our skills on the software of the brain.

What kind of software does the brain use? The brain does not work like a computer; it is not a digital device; it is an analog device. The majority of computations in the brain are performed in analog format, in the form of graded receptor and synaptic potentials, not all-or-none action potentials that, after all, end up inducing other grade potentials. Even groups of neurons, entire modules, and multimodule systems all generate waveforms of activity, from the 40 Hz rhythm thought to underlie binding of sensory events to slow potentials that may underlie long-term processes. Before we can ever hope to implant or drive machines at the macro, micro, or nano scale, the sciences of information technology and advanced computing need to sharpen our skills at analog computing. This should be the charge to our information technology colleagues: Develop analog computational software. However, we do not have to wait until we make breakthroughs in that direction, because we can go ahead and develop nanoscale *peripherals* in the meantime.

Improving Human Performance

Sensory Gating

Human performance, being under direct control from the brain, is dependent on a pyramid of processes. Accurate human performance depends on practice gained from learning and memory, which in turn depends on selective attention to the performance of the task at hand, which in turn depends on "preattentional" arousal mechanisms that determine a level of attention (e.g., I need to be awake in order to pay attention). Human performance can be improved with training, which involves higher-level processes such as learning and memory. However, the most common factor leading to poor human performance is a lower-level process, lack of attention, or distractibility. Distractibility can result from fatigue, stress, and disease, to name a few. Is it possible to decrease the degree of distractibility, or at least to monitor the level of distractibility? Can nanotechnology provide a critical service in the crucial area of distractibility?

The National Research Council's Committee on Space Biology and Medicine (1998) has concluded,

> Cumulative stress has certain reliable effects, including psychophysiological changes related to alterations in the sympathetic-adrenal-medullary system and the hypothalamic-pituitary-adrenal axis (hormonal secretions, muscle tension, heart and respiration rate, gastrointestinal symptoms), subjective discomfort (anxiety; depression; changes in sleeping, eating and hygiene), interpersonal friction, and impairment of sustained cognitive functioning. *The person's appraisal of a feature of the environment as stressful and the extent to which he or she can cope with it are often more important than the objective characteristics of the threat*.

It is therefore critical to develop a *method for measuring our susceptibility under stress to respond inappropriately to features of the environment*. "Sensory gating" has been conceptualized as a critical function of the central nervous system to filter out extraneous background information and to focus attention on newer, more salient stimuli. By monitoring our sensory gating capability, our ability to appraise and filter out unwanted stimuli can be assessed, and the chances of successful *subsequent* task performance can be determined.

One proposed measure of sensory gating capability is the P50 potential. The P50 potential is a midlatency auditory evoked potential that is (a) rapidly habituating, (b) sleep state-dependent, and (c) generated in part by cholinergic elements of the Reticular Activating System (the RAS modulates sleep-wake states, arousal, and fight versus flight responses). Using a paired stimulus paradigm, sensory gating of the P50 potential has been found to be reduced in such disorders as anxiety disorder (especially post-traumatic stress disorder, PTSD), depression, and schizophrenia (Garcia-Rill 1997). Another "preattentional" measure, the startle response, could be used, however, due to its marked habituation, measurement time is too prolonged (>20 min), and because compliance using startling, loud stimuli could also be a problem, the use of the P50 potential is preferable.

Sensory gating deficits can be induced by stress and thus represent a serious impediment to proper performance under complex operational demands. *We propose the development of a nanoscale module designed for the use of the P50 potential as a measure of sensory gating (Figure C.9).*

A method to assess performance readiness could be used as a predictor of performance success, especially if it were noninvasive, reliable, and not time-consuming. If stress or other factors have produced decreased sensory gating, then remedial actions could be instituted to restore sensory gating to acceptable levels, e.g., coping strategies, relaxation techniques, pharmacotherapy. It should be noted that this technique also may be useful in detecting slowly developing (as a result of cumulative stress) chronic sensory gating deficits that could arise from clinical depression or anxiety disorder, in which case remedial actions may require psychopharmacological intervention with, for example, anxiolytics or antidepressants.

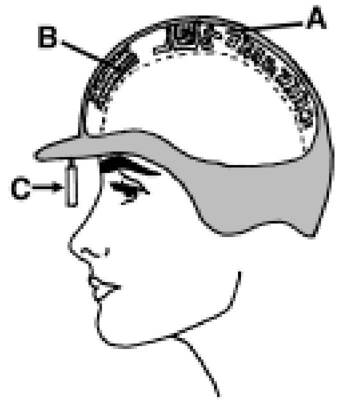


Figure C.9. Nanotechnology application: helmet incorporating P50 midlatency auditory evoked potential recording and near-infrared detection of frontal lobe blood flow to measure sensory gating and hypofrontality, respectively. A. Evoked potential module including audio stimulator (earphones), surface electrodes (vertex, mastoids, forehead), amplifiers, averager with wave recognition software, and data storage device for downloading. B. Near-infrared detection module for frontal lobe blood flow measurement. C. Flip-down screen for tracking eye movements and display of results from sensory gating and frontal blood flow measurements.

Implementation of this methodology would be limited to the ability to record midlatency auditory evoked responses in varied environments. The foreseen method of implementation would involve the use of an electronically shielded helmet (Figure C.9) containing the following: (1) P50 potential recording electrodes at the vertex, mastoids, and ground; (2) eye movement recording using a flip-down transparent screen to monitor the movements of one eye within acceptable limits that do not interfere with P50 potential acquisition; and (3) electrodes on the forehead to monitor muscle contractions that could interfere with P50 potential acquisition. The helmet would incorporate an audio stimulator for delivering click stimuli, operational amplifiers for the three measures, averaging software, wave detection software (not currently available), and simple computation and display on the flip-down screen of sensory gating as a percent. A high percentage compared to control conditions would be indicative of a lack of sensory gating (indicating increased distractibility, uncontrolled anxiety, etc.). An individual could don the helmet and obtain a measure of sensory gating *within 5-7 minutes.*

The applications for this nanotechnology would be considerable, including military uses for self-monitoring human performance in advance of and during critical maneuvers; for self-monitoring by astronauts on long-duration space missions; for pilots, drivers and operators of sensitive and complex equipment, etc. It should be noted that this physiological measure can not be “faked” and is applicable across languages and cultures.

Hypofrontality

In general, the role of the frontal cortex is to control, through inhibition, those old parts of the brain we inherited from our early ancestors, the emotional brainstem (Damasio 1999). If the frontal cortex loses some of its inhibitory power, "primordial" behaviors are released. This can occur when the cortex suffers from decreased blood flow, known as "hypofrontality." Instinctive behaviors then can be released, including, in the extreme, exaggerated fight versus flight responses to misperceived threats, i.e., violent behavior in an attempt to attack or flee. "Hypofrontality" is evident in such disorders as schizophrenia, PTSD, and depression, as well as in neurodegenerative disorders like Alzheimer's and Huntington's diseases. Decreased frontal lobe blood flow can be induced by alcohol. Damage, decreased uptake of glucose, reduced blood flow, and reduced function have all been observed in the frontal cortex of violent individuals and murderers.

The proposed method described below could be used to detect preclinical dysfunction (i.e., could be used to *screen and select* crews for military or space travel operations); to determine individual performance under stress (i.e., could be used to *prospectively evaluate* individual performance in flight simulation/virtual emergency conditions); and to monitor the effects of chronic stressors (i.e., monitor sensory gating during long-duration missions). This nanomethodology would be virtually realtime; would not require invasive measures (such as sampling blood levels of cortisol, which are difficult to carry out accurately, are variable and delayed rather than predictive); and would be more reliable than, for example, urine cortisol levels (which would be delayed, or could be compensated for during chronic stress). Training in individual and communal coping strategies is crucial for alleviating some of the sequelae of chronic stress, and the degree of effectiveness of these strategies could be *quantitatively assessed* using sensory gating of the P50 potential as well as frontal lobe blood flow. That is, these measures could be used to determine the efficacy of any therapeutic strategy, i.e., to measure outcome.

A detecting module located over frontal areas with a display on the flip-down screen could be incorporated in the helmet to provide a noninvasive measure of frontal lobe blood flow for self-monitoring in advance of critical maneuvers. The potential nanotechnology involved in such measures has already been addressed (Chance and Kang n.d.). Briefly, since hemoglobin is a strong absorber, changes in this molecule could be monitored using near-infrared detection. This promising field has the potential for monitoring changes in blood flow as well as hemoglobin saturation, a measure of energy usage.

Peripheral nanotechnology applications such as P50 potential recordings and frontal blood flow measures are likely to provide proximal, efficient, and useful improvements in human performance. Nanotechnology, by being transparently integrated into our executive functions, will become part of the enculturation process, modulating brain structure and driving our evolution.

References

Chance, B., Kang, K. 2002. Optical identification of cognitive state. *Converging technology (NBIC) for improving human performance* (this volume).

Damasio, A. 1999. *The Feeling of What Happens, Body and Emotion in the Making of Consciousness*, Harcourt Brace & Co., New York, NY.

Donald, M.W. 1991. *Origins of the Modern Mind, Three Stages in the Evolution of Culture and Cognition*, Harvard University Press, Cambridge, MA.

Garcia-Rill, E. 1997. Disorders of the Reticular Activating System. *Med. Hypoth.* 49, 379-387.

National Research Council Committee on Space Biology and Medicine. 1998. *Strategy for Research in Space Biology and Medicine into the Next Century*. National Academy Press: Washington, DC.

Pollack, J. 2002. The limits of design complexity. *Converging technology (NBIC) for improving human performance* (this volume).

SCIENCE AND TECHNOLOGY AND THE TRIPLE D (DISEASE, DISABILITY, DEFECT)

Gregor Wolbring, University of Calgary

Science and technology (S&T) have had throughout history — and will have in the future — positive and negative consequences for humankind. S&T is not developed and used in a value neutral environment. S&T activity is the result of human activity imbued with intention and purpose and embodying the perspectives, purposes, prejudice and particular objectives of any given society in which the research takes place. S&T is developed within the cultural, economical, ethical, and moral framework of the society in which the research takes place. Furthermore, the results of S&T are used in many different societies reflecting many different cultural, economical, ethical, moral frameworks. I will focus on the field of Bio/Gene/Nanomedicine. The development of Bio/Gene/Nanotechnology is — among other things — justified with the argument that it holds the promises to fix or help to fix perceived disabilities, impairments, diseases and defects and to diminish suffering. But who decides what is a disability, disease, an impairment and a 'defect' in need of fixing? Who decides what the mode of fixing (medical or societal) should be, and who decides what is suffering? How will these developments affect societal structures?

Perception

The right answers to these questions will help ensure that these technologies will enhance human life creatively, rather than locking us into the prejudices and misconceptions of the past. Consider the following examples of blatant insensitivity:

> Fortunately the Air Dri-Goat features a patented goat-like outer sole for increased traction so you can taunt mortal injury without actually experiencing it. Right about now you're probably asking yourself "How can a trail running shoe with an outer sole designed like a goat's hoof help me avoid compressing my spinal cord into a Slinky on the side of some unsuspecting conifer, thereby rendering me a drooling, misshapen non- extreme-trail-running husk of my former self, forced to roam the earth in a motorized wheelchair with my name embossed on one of those cute little license plates you get at carnivals or state fairs, fastened to the back?" (Nike advertisement, Backpacker Magazine, October 2000).
>
> Is it more likely for such children to fall behind in society or will they through such afflictions develop the strengths of character and fortitude that lead to the head of their packs? Here I'm afraid that the word handicap cannot escape its true definition — being placed at a disadvantage. From this perspective seeing the bright side of being handicapped is like praising the virtues of extreme poverty. To be sure, there are many individuals who rise out of its inherently degrading states. But we perhaps most realistically should see it as the major origin of asocial behavior (Watson 1996).

American bioethicist Arthur Caplan said in regards to human genetic technology, "the understanding that our society or others have of the concept of health, disease and normality will play a key role in shaping the application of emerging knowledge about human genetics" (Caplan 1992). I would add Nanomedicine/Nanotechnology into Caplan's quote because parts of nanotechnology development are inherently linked with bio/genetechnology as the following quote from a recent report on its societal implications illustrates:

> Recent insights into the uses of nanofabricated devices and systems suggest that today's laborious process of genome sequencing and detecting the genes' expression can be made dramatically more efficient through use of nanofabricated surfaces and devices. Expanding our ability to characterize an individual's genetic makeup will revolutionize diagnostics and therapeutics (Roco and Bainbridge 2001).

In addition, nanomedicine and nanotechnologies must be added, to quote the report again, because they

> ...hold promise for contributing to a wide range of assistive solutions, from prosthetic limbs that adjust to the changes in the body, to more biocompatible implants, to artificial retinas or ears. Other opportunities lie in the area of neural prosthesis and the "spinal patch," a device envisioned to repair damage from spinal injuries (Roco and Bainbridge 2001).

Any of these solutions are linked to the normalcy concept, the ability concept, and to the perceptions of what needs to be assisted. Certainly, different responses will be made and different solutions will be sought depending on how the problem is defined, and how the problem will be defined depends on our concepts of and beliefs about such things as health, disease, disability, impairment, and defect. For example, whether being gay is seen as a disease and defect (medical model) or a variation of human diversity (social model) will lead to totally different intervention scenarios (medical cure versus social cure). In the same way, what if we would view women as a double X syndrome, or men as an XY syndrome?

In essence every biological reality can be shaped and seen as a defect, as a medical problem, or as a human rights and social problem. No one views nowadays — in western culture at least — the biological reality of being a women within a medical framework, although a women was still viewed at the end of last century in countries like the UK as too biologically fragile and emotional and thus too dependent, to bear the responsibility attached to voting, owning property, and retaining custody of their own children (Silvers et. al., 1998). Therefore, a societal cure of equal rights and respect is seen as the appropriate remedy for the existing disparity between women and men. Gays, lesbians, bisexuals, and other groups demand that their problems are seen within a social framework and not within a medical framework.

So what now about so-called disabled people? Are "disabled people" or differently said "people who do not fit society's expectation of normal ability" to be seen as a medical problem or as part of the diversity of humankind? Within the medical model, disability is viewed as a defect, a problem inherent in the person, directly caused by disease, trauma, or other health condition and a deviation from certain norms. Management of the disability of the disabled person or person-to-be is aimed at cure, prevention, or adaptation of the person (e.g. assistive devices). Medical care and rehabilitation are viewed as the primary issues, and at the political level, the principal response is that of modifying or reforming health care policy.

The social model of disability on the other hand, sees the issue mainly as a socially created problem and principally as a matter of the full integration of individuals into society. Disability is not an attribute of an individual, but rather a complex collection of conditions, many of which are created by the environment, particularly the social environment and socially mediated aspects of the physical environment. Hence, the management of the problem requires social action, and it is the collective responsibility of society at large to make the environmental modifications necessary for the full participation of people with disabilities in all areas of social life. The issue is therefore an attitudinal or ideological one requiring social change, which at the political level becomes a question of human rights to be seen in the same way as the issue of gender and sexual orientation. In essence able-ism is seen in the same light as racism, sexism, age-ism, homophobia, etc.

The social model of disability does not negate that a disabled person has a certain biological reality (like having no legs) which makes her/him different in her/his abilities, which make her/him not fit the norm. But it views the “need to fit a norm” as the disability and questions whether many deviations from the norm need a medical solution (adherence to the norm) or a social solution (change/elimination of norm).

Many bio/gene/nano technology applications (predictive testing, cures, adaptation) focus on the individual and his or her perceived shortcomings. They follow a medical, not a social evaluation of a characteristic (biological reality) and therefore offer only medical solutions (prevention or cure/adaptation) and no social solutions (acceptance, societal cures of equal rights and respect).

Furthermore the use and development focus of bio/gene/nanotechnology as it is perpetuates the medical, intrinsic, individualistic, defect view of disability. Not often discussed by clinicians, academics in general, or the general public is the view, commonly expressed by disabled people, that the demand for the technology is based too much on the medical model of disability and hardly acknowledges the social model of disability (Asch 1999, Miringoff 1991; Hubbard 1990: Lippman 1991; Field 1993; Fine & Asch 1982; Minden 1984; Finger 1987; Kaplan 1994; Asch 1989; Asch and Geller 1996).

The perception of disabled people as suffering entities with a poor quality of life, in need of cure and fixing for the most part does not fit with the perceptions disabled people have of themselves. This fact is illustrated by Table C.5, which compares self esteem of people having spinal cord injury with the images many nondisabled people have of what this hypothetically would mean for themselves.

Table C.5: Self-esteem ratings following severe spinal cord injury (SCI)

Percent agreeing with each statement	**Nondisabled Respondents**	**Nondisabled Respondents Imagining Self with SCI**	**SCI Survivors Comparison Group**
I feel that I am a person of worth.	98%	55%	95%
I feel that I have a number of good qualities.	98%	81%	98%
I take a positive attitude.	96%	57%	91%
I am satisfied with myself on the whole.	95%	39%	72%
I am inclined to feel that I am a failure.	5%	27%	9%
I feel that I do not have much to be proud of.	6%	33%	12%
I feel useless at times.	50%	91%	73%
At times I feel I am no good at all.	26%	83%	39%

Clearly, most people with spinal cord injury have positive self-images, but nondisabled people have the false impression that a person with this injury would lack self-esteem. This table was adapted from Gerhart et al., 1994, but many other studies report similar findings (Cameron 1973; Woodrich and Patterson 1983; Ray and West 1984; Stensman 1985; Bach and Tilton 1994; Cushman and Dijkers 1990; Whiteneck et al. 1985; Eisenberg and Saltz 1991; Saigal et al. 1996 Tyson and Broyles 1996; Cooley et al. 1990).

The following passage provides an example of how many professionals view the effects of people with disabilities on their families.

> How did parents endure the shock [the birth of a thalidomide baby]? The few who made it through without enormous collateral damage to their lives had to summon up the same enormous reserves of courage and devotion that are necessary to all parents of children with special needs and disabilities; then, perhaps, they needed still more courage, because of the special, peculiar horror that the sight of their children produced in even the most compassionate. Society does not reward such courage… because those parents experience represents our own worst nightmare, ever since we first imagined becoming parents ourselves. The impact upon the brothers and sisters of the newborn was no less horrific. This was the defining ordeal of their family life — leaving aside for now the crushing burden on their financial resources from now on (Stephens and Brynner 2001).

While such negative views of the impact of children with disabilities on their families have dominated clinical and research literature for decades, more recent research has exposed these negative biases as empirically unsupportable and clinically destructive (e.g., Helf and Glidden, 1998; Sobsey, 1990). Contemporary research suggests that parents, like people with disabilities, do not view their children with disabilities as their "worst nightmares," as sources of "peculiar horror" or as "crushing burdens." In fact, most view them very much as they view children without disabilities, as sources of significant demands but even greater rewards (e.g., Sobsey & Scorgie 2001). Yet, people with disabilities and their families are a part of society and they can never be entirely free of the attitudes, beliefs, and biases held by professionals and the general public.

Such attitudes and beliefs about disability contribute to the drive to fix people with disabilities rather than accommodate them. For example, the quote from Stephens and Brynner seems to suggest:

1. an implicit assumption of normalcy which requires two legs and two arms
2. an expectation that everyone has to be able to perform certain functions (e.g., move from one place to another or eat)
3. an expectation that everyone has to perform this function in a the same way (e.g., walking upright on their own legs or eat with their hands)
4. an expectation that any variation in form, function, method will result in severe emotional distress for those involved in any way

These attitudes drive the development of artificial legs and arms and help to explain why thalidomide kids and their parents were confronted with the single-minded approach to outfit thalidomide kids with artificial limbs without exploring different forms of functioning. Professionals typically persisted with this approach in spite of the fact that artificial limbs were rather crude, not very functional, and mostly cosmetic at the time and they were being prescribed in great numbers. The approach nearly completely excluded alternatives, such as crawling in the absence of legs or eating with one's feet in the absence of arms. The sentiment expressed by Stephens and Brynner also prevents adaptation by society to alternative modes of function (e.g., moving and eating).

This kind of single-minded approach reflects an adherence to a certain norm, which was more readily accepted by amputees who lost their arms or legs. They were or are willing to accept this because in a large part due to the fact that they were not allowed to adapt and get used to their new condition, a process that we all know takes time. People take time to adapt to any change. Humankind is not known for its ability to adapt easily to changes (e.g., divorce, career changes). Thalidomiders did not have to readapt to a new body reality. That might explain why most Thalidomiders threw away their artificial legs and arms as soon as they were old enough to assert themselves against their parents and the medical profession. For them the reality was that they did not view their body as deficient and did not see artificial legs or arms as the most suitable mode of action. In light of the perception reflected in the Stephens and Brynner's quote, the question becomes whether the development of usable artificial legs and arms mean that someone without legs or arms will be even more stigmatized if he or she does no use them. If so, the presence of this option is not merely another free choice since existence of the option results in a coercive influence on those who might refuse it.

Choice

The question arises whether usable artificial limbs increase choice as an optional tool or establish a norm that restricts choice. Parents of Thalidomiders were not given much choice. Immense pressure was used to have the parents equip their kids with artificial limbs. Society already judges certain tools. A hierarchy regarding movement exists. Crawling is on the bottom of the acceptance list, below the wheelchair, which is seen as inferior to the artificial leg particularly one that appears "natural." This hierarchy is not based on functionality for the person but rather on emotions, prejudice, and rigid adherence to a normative body movement. Tools like the wheelchair are frequently demonized in expressions such as "confined to the wheelchair." It is interesting that people do not say "confined to" artificial legs even though a wheelchair often leads to safer, easier, and more efficient mobility for an individual than artificial legs do. No one would use the phrase "confined to natural legs" for "normal" people, although in reality they are confined to their legs while many wheelchair users can leave their wheelchairs. Similarly, the negative concept of confinement is not used to describe driving a car, which is viewed as empowering rather than limiting, even though many of us are heavily dependent on this mode of transportation. In much the same way, most of us who live in the north would not survive a single winter without central heating but we generally do not label all of these people as "technology dependent."

Cochlear implants provide another related example. Do we allow parents to say "No" to them if they feel there is nothing wrong with their kid using sign language, lip reading, or other alternative modes of hearing? Will the refusal by the parents be viewed as child abuse (see Harris, 2000 for an ethical argument to view it as child abuse)? Might parents have been considered to commit child abuse if they had refused artificial limbs for their thalidomide kids? Or in today's world, could a mother be considered to commit child abuse if she refused to terminate her pregnancy after ultrasound showed phocomelia (i.e., hands and feet attached close to the body without arms or legs) in the fetus. Of course, ultrasound wasn't an option when most of the thalidomide cases occurred but it is today. Furthermore, would the mother abuse society by not fixing (cure, adaptation, prevention) the "problem"?

A hint to the answer to these questions is given by the following results of a survey of genetic counselors in different countries (Wertz 1998):

> The majority in 24 countries believed it is unfair to the child to be born with a disability. 40% agreed in USA, Canada and Chile. 36% in Finland and UK; 33% in Switzerland and the Netherlands; 29% in Argentina, 27% in Australia 25% in Sweden and 18% in Japan.
>
> It is socially irresponsible knowingly to bring an infant with a serious [no legal document defines what is serious] genetic disorder into the world in an era of prenatal diagnosis." More than 50% agreed in South Africa, Belgium, Greece, Portugal, Czech Republic, Hungary, Poland, Russia, Israel, Turkey, China, India, Thailand, Brazil, Columbia, Cuba, Mexico, Peru and Venezuela. 26% of US geneticists, 55% of US primary care physicians and 44% of US patients agreed.

> A high percentage of genetic counselors feels that societies will never provide enough support for people with disabilities. The percentage of agreement for the statement ranges from 18% as a lowest to 80% in the UK. Germany is in the middle with 57%. The USA has a number of 65%.

These statements suggest that women don't have a free choice but to are led to follow the path of medical intervention. In the absence of a possible social cure for disability, the only option left that may appear to be available is the medical cure in whatever shape and form, independent of its usefulness and need.

The treatment of Thalidomiders, the pressure to install cochlear implants, and prebirth counseling raise a more general question about whether advances in a wide range of assistive devices, partly due to advances in micro- and nanotechnologies, will lead to increased or restricted choices. We can hope that technological convergence offers humanity so many choices that false stereotypes about the disabled are discredited once and for all. But this can happen only if we recognize the alternatives as real choices that must be considered with sensitivity, imagination, and — most importantly — the judgment of disabled people themselves.

Consequences

The history of the debate around bio/gene/nano-technology as it relates to disability shows a strong bias towards a medical, individualistic, intrinsic defect view of disability focusing on medical/technological cures without addressing societal components. People who promote the use of bio/genetechnology often denounce the social model of disability (Harris 2000; Singer 2001).

The medical model of disability can also show itself in court rulings, such as some recent US Supreme Court rulings. The Supreme Court ruled on the "definition of disability" in Sutton v. United Airlines (130 F.3d 893, 119 S. Ct. 2139), Albertsons Inc. v. Kirkingburg (143 F.3d 1228, 119 S. Ct. 2162), and Murphy v. United Parcel (141 F.3d 1185, 119 S. Ct. 1331), stating that the Americans with Disabilites Act does not cover those persons with correctable impairments.[1] In other words, as soon as adaptations are available, all problems must be fixed and no protections through civil rights laws, such as the ADA, are allowed anymore. Not only that the ruling implies that disability is something which can be fixed through medical technological means. A social view of disability does not fit with the above ruling.

We see a disenfranchisement of disabled people from the equality/human rights movement. (Wolbring 1999, 2000, and 2001). So far, bio/genetechnology has led to an increase in discrimination against characteristics labeled as disabilities, as the following three examples illustrate.

First, we see a proliferation of legal cases involving wrongful life or wrongful birth suits (Wolbring, 2001,2002a). Wrongful life suits are only accepted if the child is disabled. And wrongful birth suits are specific by now for disability with special rulings whereas cases based on non-disability are called wrongful pregnancy. The remedies in the case of wrongful birth/pregnancy cases are quite different. The following quotations illustrate the logic of such cases.

> Two other justices based their agreement of wrongful life suits on the view that the physicians wrongful life liability towards the disabled infant as resting on the right to life without a handicap. Thus the damage is measured by comparing the actual impaired life of

[1] National Council on Disability USA, 2000; Civil Rights, Sutton v. United Airlines, Albertsons Inc. v. Kirkingburg, and Murphy v. United Parcel (http://www.ncd.gov/newsroom/publications/policy98-99.html#1).

> the plaintiff to a hypothetical unimpaired life (CA 518, 540, 82 Zeitzoff versus Katz (1986) 40 (2) PD 85 Supreme Court of Israel (482); Shapiro 1998).
>
> ...in essence … that [defendants] through their negligence, [have] forced upon [the child] the worse of … two alternatives, … that nonexistence — never being born — would have been preferable to existence in the diseased state (Soeck v. Finegold, 408 A.2d 496(Pa. 1970)).
>
> Thus the legislature has recognized," the judge said, "as do most reasonable people, that cases exist where it is in the interest of the parents, family and possible society that it is better not to allow a fetus to develop into a seriously defective person causing serious financial and emotional problems to those who are responsible for such person's maintenance and well-being (Strauss 1996).

Second, anti-genetic discrimination laws cover discrimination on genetic characteristics which might lead in the future to 'disabilities' in a medical sense but are for the time being asymptomatic. In essence, the feature of genetic discrimination is the use of genetic information about an asymptomatic disabled person. The vogue for the establishment of an Anti-Genetic Discrimination law for asymptomatic disabled people highlights one other reality, namely that symptomatic disabled people are excluded from exactly the benefits the Anti-Genetic Discrimination laws try to address. With these new laws these symptomatic disabled people will still be discriminated against whereas the asymptomatic ones will be safe. Not only that, ability becomes a measure to justify these new laws, as the following statement from the American Civil Liberties Union illustrates.

> The ACLU believes that Congress should take immediate steps to protect genetic privacy for three reasons. First, it is inherently unfair to discriminate against someone based on immutable characteristics that do not limit their abilities... (ACLU 2000)
>
> In sum, the ACLU believes that Americans should be judged on their actual abilities, not their potential disabilities. No American should lose a job or an insurance policy based on his or her genetic predisposition. (ACLU 2000)

A third consequence of the current mindset is differential use of genetic predictive testing. We see an Animal Farm Philosophy in regards to what to test for. Testing to eliminate any so called disability, disease, defect is acceptable but testing to determine and select on the basis of a characteristic like sex is not (Wolbring 2000, 2001).

Where should we go from here? To prevent further stigmatization, recommendations such as those quoted below from the UNESCO World Conference on Sciences 1999 conference should be implemented.

> 25. ...that there are barriers which have precluded the full participation of other groups, of both sexes, including disabled people, indigenous peoples and ethnic minorities, hereafter referred to as "disadvantaged groups..."
>
> 42. Equality in access to science is not only a social and ethical requirement for human development, but also a necessity for realizing the full potential of scientific communities worldwide and for orienting scientific progress towards meeting the needs of humankind. The difficulties encountered by women, constituting over half of the population in the world, in entering, pursuing and advancing in a career in the sciences and in participating in decision-making in science and technology should be addressed urgently. There is an equally urgent need to address the difficulties faced by disadvantaged groups, which preclude their full and effective participation.

Thus, it is essential that the greatest possible diversity of people participate in the development of convergent technologies and contribute to the associated sciences:

17. Scientists, research institutions and learned scientific societies and other relevant non-governmental organizations should commit themselves to increased international collaboration including exchange of knowledge and expertise. Initiatives to facilitate access to scientific information sources by scientists and institutions in the developing countries should be especially encouraged and supported. Initiatives to fully incorporate women scientists and other disadvantaged groups from the South and North into scientific networks should be implemented. In this context efforts should be made to ensure that results of publicly funded research will be made accessible.

79. The full participation of disadvantaged groups in all aspects of research activities, including the development of policy, also needs to be ensured.

81. Governments and educational institutions should identify and eliminate, from the early learning stages on educational practices that have a discriminatory effect, so as to increase the successful participation in science of individuals from all sectors of society, including disadvantaged groups.

91. Special efforts also need to be made to ensure the full participation of disadvantaged groups in science and technology, such efforts to include:

removing barriers in the education system;

removing barriers in the research system;

raising awareness of the contribution of these groups to

science and technology in order to

overcome existing stereotypes;

undertaking research, supported by the collection of data,

documenting constraints;

monitoring implementation and documenting best practices;

ensuring representation in policy-making bodies and forums (UNESCO 2000)

We should strive to eliminate able-ism and promote the acceptance of diversity in abilities for the sake of humankind as the best defense against gene-ism, which might affect 60% of society according to a New Zealand study. This acceptance of diverse abilities is actually also needed for the thriving of assistive technologies. For example, if an assistive technology leads to better vision than humankind has normally, should we discard the now majority of people who are less able? Or should we force all to use the new adaptive devices? Or should we demonize the ones who are more able?

The labeling of people and groups within a medical disease defect model against their will is unacceptable. In essence every scientist whose work has societal consequences has to become a societal activist to prevent these consequences.

Conclusion

The views expressed here are not opposed to progress in science and technology. As a lab bench biochemist, it would be strange for me to oppose S&T in general. Rather, this essay emphasizes the importance of openness to different perspectives on what qualifies as progress (Wolbring, 2002b). Science and Technology can be extremely useful, but certain perceptions, stereotypes, and societal dynamics can lead scientists and engineers to focus on certain types of S&T, quite apart from their objective utility to potential users.

This is not merely an issue of fairness to diverse groups of people, including the disabled. It is also an issue of imagination and insight. Convergent technologies will accomplish so much more for humanity, and unification of science will lead to so much greater knowledge, if they are free of the ignorant prejudices of the past. Specifically, science and engineering will benefit from the varied perspectives that the disabled may have about what it means to improve human performance. One essential tool to achieve this is to make sure that the teams of researchers, designers, and policy makers include many talented people who happen to be disabled.

References

ACLU. 2000. http://www.aclu.org/congress/rightgenetics.html and July 19, 2000, Letter to The Honorable Edward M. Kennedy. Ranking Member Committee on Health, Education, Labor & Pensions. 428 Dirksen Senate Office Building. Washington, D.C. 20510. http://www.aclu.org/congress/l071900a.html.

Asch, A. 1999. Prenatal diagnosis and selective abortion: A challenge to practice and policy. *American Journal of Public Health*. 89, 11, 1649.

Asch A. 1989. Reproductive technology and disability. In Cohen S, Taub N. *Reproductive laws for the 1990s.* Clifton, NJ: Humana Press. 69_124.

Asch A, Geller G. 1996. Feminism, bioethics and genetics. In: Wolf S, ed. *Feminism and bioethics: Beyond reproduction.* New York, NY. Oxford University Press. 318-350.

Bach, JR and Tilton. MC 1994. Life satisfaction and well-being measures in ventilator assisted individuals with traumatic tetraplegia. *Archives of Physical Medicine and Rehabilitation*. vol. 75, 626-632.

Cameron P et al. 1973. The life satisfaction of nonnormal persons, *Journal of Consulting and Clinical Psychology*. vol. 41, 207-214.

Caplan, Arthur L. 1992. If gene therapy is the cure, what is the disease? in Annas, George J.; Elias, Sherman, eds. *Gene mapping: Using law and ethics as guides*. New York. Oxford University Press. page 128-14.

Cooley, WC et al. 1990, Reactions of mothers and medical profession to a film about Down Syndrome *Am. J. Dis. Child.* 144 1112.

Cushman, LA and Dijkers, MP. 1990. Depressed mood in spinal cord injured patients: staff perceptions and patient realities. *Archives of Physical Medicine and Rehabilitation*. vol. 71. 191-196.

Eisenberg MG and Saltz. CC 1991. Quality of life among aging spinal cord injured persons: long term rehabilitation outcomes. *Paraplegia.* vol. 29, 514-520.

Field NIA. 1993. Killing “the handicapped” before and after birth. *Harvard Women’s Law J* 16:79-138.

Fine M, Asch A. 1982. The question of disability: no easy answers for the women’s movement. *Reproductive Rights Newsletter*. 4(3). 19-20.

Finger, A. 1987. *Past due: Disability, pregnancy and birth*. Seattle. Washington Seal Press.

Kaplan D. 1994. Prenatal screening and diagnosis: the impact on persons with disabilities. In Rosenberg KHZ, Thompson JED, eds. *Women and prenatal testing: Facing the challenges of genetic technology.* Columbus. Ohio State University Press. 49-61.

Gerhart, KA et al. 1994. Quality of life following spinal cord injury. Knowledge and attitudes of emergency care providers. *Annals of Emergency Medicine*. vol. 23. 807-812.

Harris, J. 2000. Is there a coherent social conception of disability? *J. of Medical Ethics* 26. pp. 95-100.

Helf, C. M., & Glidden, L. M. 1998. More positive or less negative? Trends in research on adjustment of families rearing children with developmental disabilities. *Mental Retardation*, 36. 457-464.

Hubbard, R. 1990. *The politics of women in biology*. New Brunswick, NJ. Rutgers University Press. chap 12-14.

Lippman, A. 1991. Prenatal genetic testing and screening: constructing needs and reinforcing inequities. *Am J Law Med.* 17(1_2): 15_50;

Minden, S. 1984. Born and unborn: the implications of reproductive technologies for people with disabilities. In: Aridity R, Duello-Klein R, Minding S, eds. *Test-tube women: What future for motherhood?* Boston, Mass. Pandora Press. 298-312.

Miringoff, ML. 1991. *The social costs of genetic welfare*. New Brunswick, NJ: Rutgers University Press.

Ray C, West J. 1984. Social, sexual and personal implications of paraplegia. *Paraplegia*. 22:75-86.

Roco, M.C., and W.S. Bainbridge (eds). 2001. *Societal Implications of Nanoscience and Nanotechnology*. Dordrecht, Netherlands: Kluwer (also available at http://nano.gov).

Saigal, S, et al.; 1996. Self-perceived health status and health-related quality of life of extremely low-birth-weight infants at adolescence. *JAMA*. 276: 453-459.

Silver, A et al. 1998. Disability, Difference, Discrimination: Perspective on justice in bioethics and public policy. Rowman & Littlefield Publishers, INC. Landham, Bolder, New York, Oxford.

Scorgie, K., & Sobsey, D. 2000. Transformational outcomes associated with parenting children with disabilities. *Mental Retardation*. 38(3), 195-206

Shapiro A1998 'Wrongful life' lawsuits for faulty genetic counselling: Should the impaired newborn be entitled to sue? *J. of Medical Ethics* 24. 369-375.

Sobsey, D. 1990. Too much stress on stress? Abuse and the family stress factor. *Quarterly Newsletter of the American Association on Mental Retardation*. 3, 2, 8.

Singer. 2001. Response to Mark Kuczewski. *American Journal of Bioethics*. Volume 1. Number 3. p. 55–56

Stensman, S, 1985. Severely mobility-disabled people assess the quality of their lives. *Scandinavian Journal of Rehabilitation Medicine*. vol. 17. 87-99.

Stephens, T and Brynner, R. 2001. *Dark Remedy; the impact of thalidomide*. Perseus Publishing. Cambridge Massachusetts. USA page 65/66.

Strauss, SA. 1996. 'Wrongful conception', 'wrongful birth' and 'wrongful life': the first South African cases. *Med. Law*. 15: 161-173.

Tyson JE, Broyles RS. 1996. Progress in assessing the long-term outcome of extremely low-birth-weight infants. *JAMA*. 276: 492-493.

UNESCO, 2000. http://unesdoc.unesco.org/images/0012/001229/122938eo.pdf#xml=http://unesdoc.unesco.org/ulis/cgi-bin/ulis.pl?database=ged&set=3BE443E4_0_108&hits_rec=3&hits_lng=eng

Watson, JD. 1996. President's essay: genes and politics. *Annual Report Cold Springs Harbor*. 1996:1-20." http://www.cshl.org/96AnReport/essay1.html. Exact page http://www.cshl.org/96AnReport/essay14.html.

Wertz, DC. 1998. Eugenics is alive and well. *Science in Context* 11. 3-4. pp 493-510 (p501).

Whiteneck, GC et al. 1985. *Rocky mountain spinal cord injury system*. Report to the National Institute of Handicapped Research. 29-33.

Wolbring. 1999. *Gene Watch* June 1999 Vol.12 No.3; http://www.bioethicsanddisability.org/Eugenics,%20Euthanics,%20Euphenics.html

Wolbring. 2000. Science and the disadvantaged; http://www.edmonds-institute.org/wolbring.html

Wolbring. 2001. Surviving in a technological world. In *Disability and the life course: Global perspectives*. Edited by Mark Priestley. Cambridge University Press.

Wolbring. 2001. 150 page single spaced expert opinion for the Study Commission on the Law and Ethics of Modern Medicine of the German Bundestag with the title "Folgen der Anwendung genetischer Diagnostik fuer behinderte Menschen" (Consequences of the application of genetic diagnostics for disabled people) http://www.bundestag.de/gremien/medi/medi_gut_wol.pdf

Wolbring. 2002a. International Center for Bioethics, Culture and Disability. http://www.bioethicsanddisability.org/wrongfulbirth.html

Wolbring. 2002b. Wrongful birth/life suits. http://www.bioethicsanddisability.org

Woodrich, W. and Patterson, JB. 1983. Variables related to acceptance of disability in persons with spinal cord injuries. *Journal of Rehabilitation*. July-Sept. 26-30.

VISIONARY PROJECTS

BRAIN-MACHINE INTERFACE VIA A NEUROVASCULAR APPROACH

Rodolfo R. Llinás and Valeri A. Makarov, NYU Medical School

The issue of brain-machine (computer) interface is, without a doubt, one of the central problems to be addressed in the next two decades when considering the role of neuroscience in modern society. Indeed, our ability to design and build new information analysis and storage systems that are sufficiently light to be easily carried by a human, will serve as a strong impetus to develop such peripherals. Ultimately, the brain-machine interface will then become the major bottleneck and stumbling block to robust and rapid communication with those devices.

So far, the interface improvements have not been as impressive as the progress in miniaturization or computational power expansion. Indeed, the limiting factor with most modern devices relates to the human interface. Buttons must be large enough to manipulate, screens wide enough to allow symbol recognition, and so on. Clearly, the only way to proceed is to establish a more direct relation between the brain and such devices, and so, the problem of the future brain-machine interface will indeed become one of the central issues of modern society. As this is being considered, another quite different revolution is being enacted by the very rapid and exciting developments of nanotechnology (n-technology). Such development deals with manufactured objects with characteristic dimensions of less than one micrometer. This issue is brought to bear here, because it is through n-technology that the brain-machine bottleneck may ultimately be resolved. Obviously, what is required is a robust and noninvasive way to both tap and address brain activity optimized for future brain-machine interaction.

Needless to say, in addition to serving as a brain-machine interface, such an approach would be extraordinarily valuable in the diagnosis and treatment of many neurological and psychiatric conditions. Here, the technology to be described will be vital in the diagnosis and treatment of abnormal brain function. Such technology would allow constant monitoring and functional imaging, as well as direct modulation of brain activity. For instance, an advanced variation of present-day deep brain stimulation will be of excellent therapeutic value. Besides, interface with “intelligent” devices would significantly improve the quality of life of disabled individuals, allowing them to be more involved in everyday activity.

The problem we consider has two main parts to be resolved: (1) hardware and (2) software. To approach these issues, we propose to develop a new technology that would allow *direct* interaction of a machine with the human brain and that would be secure and minimally invasive.

The Neurovascular Approach

One of the most attractive possibilities that come to mind in trying to solve the hardware problem concerns the development of a vascular approach. The fact that the nervous system parenchyma is totally permeated by a very rich vascular bed that supplies blood gas exchange and nurturing to the brain mass makes this space a very attractive candidate for our interface. The capillary bed consists of 25,000 meters of arterio-venous capillary connections with a gage of approximately 10 microns. At distances more proximal to the heart, the vessels increase rapidly in diameter, with a final dimension of over 20 millimeters. Concerning the acquisition of brain activity through the vascular system, the use of n-wire technology coupled with n-technology electronics seems very attractive. It would allow the nervous system to be addressed by an extraordinarily large number of isolated n-probes via the vascular bed, utilizing the catheter-type technology used extensively in medicine and in particular in interventional neuro-radiology.

The basic idea consists of a set of n-wires tethered to electronics in the main catheter such that they will spread out in a “bouquet” arrangement into a particular portion of the brain’s vascular system. Such arrangement could support a very large number of probes (in the millions). Each n-wire would be used to record, very securely, electrical activity of a single or small group of neurons without invading the brain parenchyma. Obviously, the advantage of such system is that it would not interfere with either the blood flow exchange of gases or produce any type of disruption of brain activity, due to the tiny space occupied in the vascular bed.

In order to give a more precise description of the proposed interface, an illustration of the procedure is shown in Figure C.10. A catheter is introduced into the femoral carotid or the sub-clavial artery and is pushed up to one of the vascular territories to be addressed. Such procedure is, in principle, similar to interventional neuro-radiology techniques where catheters are guided to any portion of the central nervous system. The number of 0.5 micron diameter wires (recording points) that could be introduced in a one-millimeter catheter is staggeringly large (in the range of few million). Once the areas to be recorded or stimulated are reached, a set of leads held inside the catheter head would be allowed to be extended and randomly distributed into the brain’s circulatory system. Since a catheter can be placed in any major brain vessels, the maximum length of n-wire electrodes required to reach any capillary bed is of the order 2 to 3 cm. Hence, a large number of electrodes would cover any region of the central nervous system from the parent vessels harboring the stem catheters.

General Electronic Design

A number of single n-wire electrodes can be attached via amplifier-binary converter to a multiplex amplifier that would sequentially switch between large, “simultaneously recorded” electrical brain signals (Figure C.10B). This is possible since the switching properties of modern multiple amplifiers are many orders of magnitude faster than the electrical signals of the brain. Thus, the number of independent wires necessary to convey the information down to the terminals of the interface would be a small fraction of the total number of n-wires, and thus, inexpensive and robust microwires can be used along the catheter length.

Many technical issues concerning hardware problems, such as n-amplifiers and multiplex units, can in fact be solved by present technology. The actual size of the expected extracellular recording wiring is given in Figure C.11 by comparing the size of one-micrometer wire with the size of a capillary in the brain parenchyma. In this case, an individual Purkinje cell is drawn to show where the capillary spaces reside within the dendritic tree of such neurons. Note that the number of capillaries traversing each cell is numerous (in this particular case, more then 20). On the right inset is an electron micrograph of the same area that gives an accurate representation of the size relation between one such n-wire (in this case 0.9 micron) and the diameter of the smallest of capillaries in that portion of brain parenchyma.

Thus, at this point, the materials and methodology required to implement a mega electrode system are basically within our technology over the next decade.

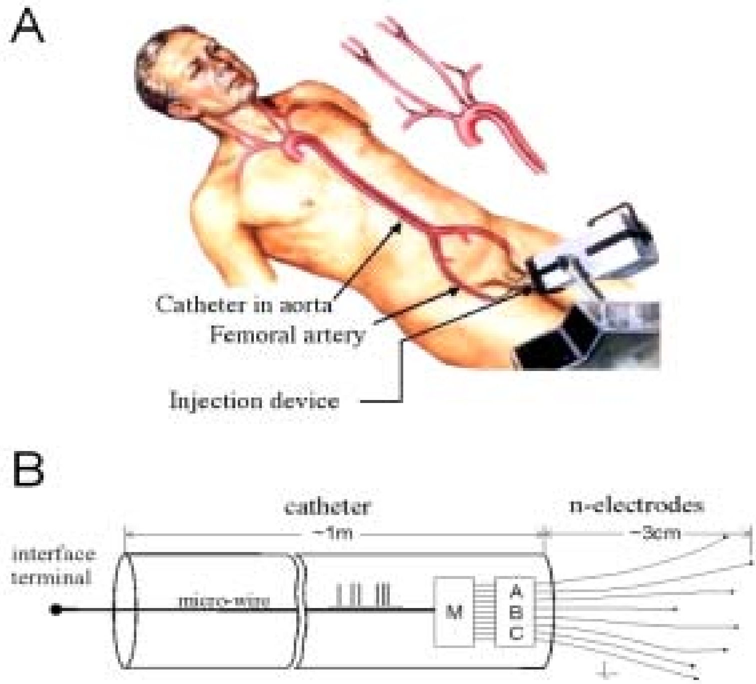


Figure C.10. The neurovascular approach. A. Present day procedure utilized to guide catheters to the brain via the vascular system. Catheters are introduced into femoral, subclavial, or carotid artery. B. The general electronic design includes n-electrodes (diameter of 0.5 micron and length not more than 3 cm) to record/stimulate neuronal activity; Amplifier-Binary Converter (ABC) block that converts acquired analog signals into binary form; Multiplex (M) unit that transforms analog input into serial form by fast switching between all signals; and microwire (approx. 1 m long) that conveys information to the terminal. (Only one logic set is shown.)

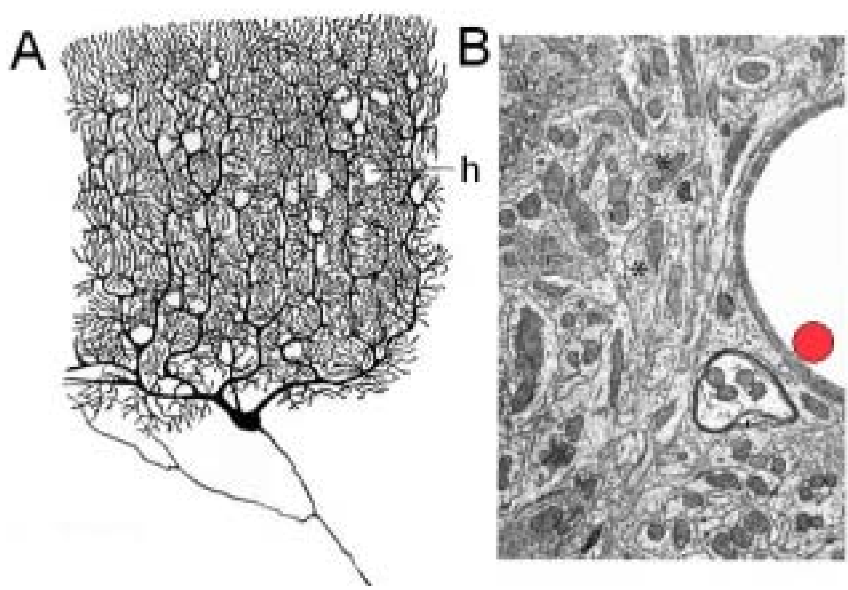


Figure C.11. Illustration of comparative size scales for a neuron, a capillary, and an n-wire. A. Purkinje cell with dendritic tree penetrated by many capillaries foramen. h. B. Elentronmicrograph of a corresponding site in the dendritic as shown in h with a 1µ electrode (spot) drawn inside a capillary.

Software Requirements

The second significant issue is that of the computational requirements that would allow the reading, storing and contextualizing of the enormous amount of neuronal information that would become available with the vascular approach described above. While this may prove to be more challenging than the hardware component of this interface, it would also be most valuable, as the proper understanding of such activity would give us an significant window into brain function, further defining the relations between electrophysiology and cognitive/motor properties of the brain.

Attempting to investigate this problem, the second step in this proposal, would be the development of mathematical algorithms able to classify brain states based on neuronal unit activity and field potential analysis. Initially, we plan to correlate, in real time, the moment-to-moment electrical activity of neurons with large functional brain states. It is assumed that the electrical properties of neurons define all possible brain states and that such states co-vary systematically with the global state dynamics. However, this does not imply that there exists one-to-one correspondence between purely local patterns of brain activity and a particular set of functional states. The generation of a new functional state in the brain, for instance, transition "sleep-wakefulness," is known to correspond to activity reorganization over many groups of neurons. Needless to say, there is a large number of possible patterns that differs minimally from one other. The approach is to map the small variance patterns into relatively small sets of different functional states. For example, in the simplest case only three global functional states may be considered: (1) sleep, (2) wakefulness, and (3) "none of the above" or uncertain state, e.g., drowsiness. The last state is an absolutely necessary form to be included, for two reasons: (a) mathematically, the output domain of the algorithm must be closed in order to address correctly "any possible input pattern," including those that have unavoidable noise impact or belong to intermediate, non-pure states without a reliable answer within statistical significance level; and (b) from the conceptual viewpoint, the third state is vital, as for instance, seeing can only occur during wakefulness, and during sleep, this state is uncertain.

The design of the hardware part of the interface (see Figure C.10B) has not been dictated by electronic purposes only but also pursues the goal of preliminary signal processing. Here, we use the commonly accepted hypothesis that neurons interact with each other mostly via action potentials and related synaptic interactions. Thus, it seems to be natural to convert electrical signals taken from n-electrodes into binary form. This approach has many advantages. In particular, if the threshold level for digitalization is appropriately chosen, we would be able to overcome the following problems:

- Not all electrodes would be placed at "right" positions (some of them may be far enough from any neuron to produce reliable data), or just damaged.
- Two electrodes placed in vicinity of a single neuron but at diverse distances from it will produce output voltage traces of different amplitude.
- The signal-to-noise ratio may not be optimal if an electrode records from more than one neuron, as one of them may be selected and others suppressed by the threshold system.

Moreover, binary form is computer friendly and supports efficient operation. Also additional processing logic can be easily included between a computer and the terminals of microwires that would significantly speed up data acquisition, storage, and contextualization.

Memory Requirements

A rough estimate of memory requirements to support resident information and input bandwidth (informational flow rate) will be $10^6 \times 10^3 = 10^9$ bits/s, assuming input signals from 10^6 independent binary variables with a sampling rate of 1 kHz. That is 100 MB per second for the total output, which is attainable with present day technologies. Utilization of additional intermediate logic would even afford a greater performance increase.

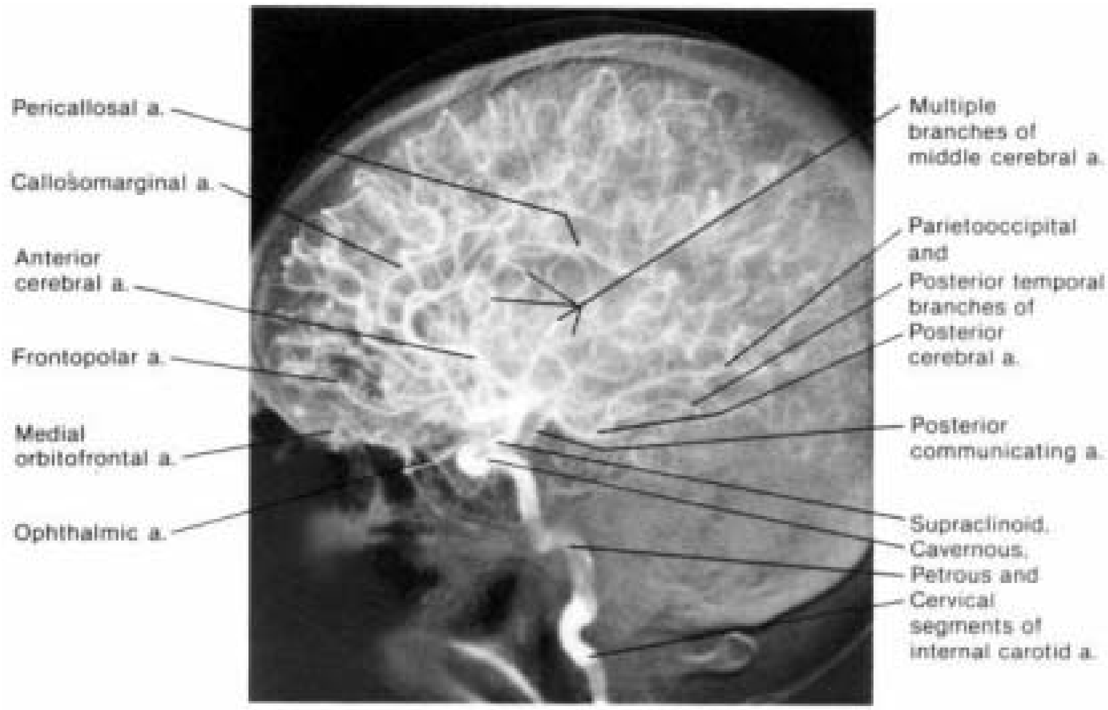


Figure C.12. Lateral view of brain arteries.

Classification Algorithms

As mentioned above, the computational algorithm must be designed to spot alterations in the brain activity that relate to a global change of states. This activity is represented by the set of binary time series taken from many neurons, i.e., by spatiotemporal patterns. Thus, we have the pattern classification problem mentioned above. For an algorithm to be useful, it must be optimized to (1) determine the minimal number of hypotheses (possible functional states) concerning the data set; (2) economize on data storage and subsequent data manipulation/calculation; (3) scale for increasing data sets and for the number of functional states; and (4) be robust. The approach to the problem we propose below is based on *cluster analysis* (Kaufman 1990) and measures of dissimilarity between time series (see, for example, Kantz 1994; Schreiber 1997; Schreiber and Schmitz 1997).

In the first step, the data set will be split into J short time intervals by shifting a time window of length T. The time scale T can be varied for different purposes, and its choice is a compromise between speed and reliability in data analysis. Each window will be referred to as "an object" or entity, assuming that a window encompasses an unchanged functional state. Assuming a correct set of hypotheses concerning the number of clusters, K, (e.g., for three global functional states: wakefulness, sleep, and uncertain state, K=3), the J different objects must be related to K functional states.

The algorithm starts with K random clusters and then moves objects between those clusters in order to split objects into clusters such that variance in each cluster would be minimal, while variance between clusters would be maximal. This can be realized by minimization of the so-called cost function (Schreiber and Schmitz 1997). To implement this function, a measure of dissimilarity between objects must be obtained. This can be, for instance, determined by calculating Euclidean distances between objects in a multidimensional space. Figure C.13 shows a sketch of average dissimilarity of object *j* to cluster *k* (distance between *j* and *k*) and average dissimilarity within cluster *k*. The optimization strategy to determine the absolute minimum of the cost function will employ simulated annealing (Kirkpatrick, Gelatt, and Vecchi 1983; Press et al. n.d.), which follows local gradients, but can move against the gradient in order to escape "local minima" shadowing an absolute minima.

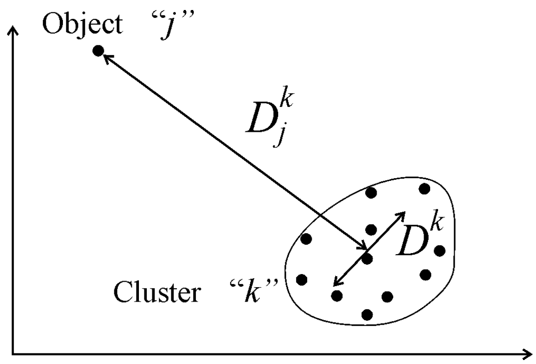


Figure C.13. Qualitative illustration of dissimilarity of object "j" to cluster "k" and mean dissimilarity within the cluster.

The algorithm described above works well under the assumption that the correct dissimilarity has been determined. For time series objects, in the simplest case, neuronal firing rates can be used as coordinates in a multidimensional space. However, application of this measure is rigid (although it has its own advantages), as it takes into account only local oscillatory properties. Another useful procedure will be the dissimilarity matrix calculation introduced (Schreiber and Schmitz 1997) based on the Grassberger-Procaccia cross-correlation sum (Grassberger and Procaccia 1983).

The classification algorithm given here may be referred to as unsupervised. It is based on the hypothesis of a "good" dissimilarity measure and does not include any optimization. This approach can be upgraded to a supervised training data set, where the correct results of classification are known *a priori* for a part of data and may be used as a feedback for improvement of computational speed and reliability. However, even after tuning, the algorithm may fail, since brain plasticity may occur. Thus, the possibility of sudden mistakes may be corrected by means of the feedback.

The basic problem here is the nonstationary nature of brain function. This seems at first glance to be a big obstacle for any time series analysis. However, a detailed study of the problem indicates two features: First, all functional states are temporal and have essentially different time scales. For example, being awake can last for hours, while cognition can be as short as tens of milliseconds. Second, we may assume that only a limited number of functional states can coexist. These two features allow building a new adaptive algorithm capable of discriminating, in principle, any possible functional states.

There are three main parameters at play. The first is length of the time window, *T*; the next is the number of clusters of objects, *K*, being separated; and the third is a dissimilarity measurement. We can start the process of classification with relatively long *T* and small *K*. Thus, fast processes (functional states) would be eliminated due to averaging over a protracted time. Moreover, functional states with intermediate time scales and with a strong influence onto others would be left out due to very rough classification, since we have split patterns into a few clusters only. Then, when a first approximation of cluster boundaries is determined and it can reliably detect functional states of the top level, a step down can be taken by decreasing window size *T* and by including finer functional states (increasing *K*). Moreover, it is possible to work "within" a functional state of the upper level and reject all non-fitting. Such modification of the algorithm allows scalability and a method of exploration of all possible functional states. One problem here is that the deeper we go into the functional state hierarchy, the heavier the computation needed. However, the main parts of the algorithm can be easily paralleled and hence effectively performed by parallel computers or even by specially designed electronics.

Conclusions

We proposed that a novel brain-machine interface is realizable that would allow a robust solution to this important problem. This hardware/software approach allows a direct brain interface and the classification of its functional states using a benign invasive approach. We propose that this approach would be very helpful in human capacity augmentation and will yield significant new information regarding normal and abnormal brain function. Because its development and utilization is inevitable given the extraordinarily attractive feature of being retrievable, in the sense that the recording/stimulating filaments are small enough that the device can be removed without violating the integrity of the brain parenchyma.

Because such interfaces will probably be streamlined over the coming years in efforts such as "hypervision" (Llinás and Vorontsov in preparation), two-way direct human communication, and man-machine telepresence (which would allow actuator-based distant manipulation), this approach should be fully examined. Finally, the development of new nanotechnology instrumentation may ultimately be an important tool in preventive medicine and in diagnostic/therapeutic outcome monitoring of physiological parameters.

References

Grassberger, P. and I. Procaccia. 1983. *Physica (Amstredam)* D 9, 189 (1983).

Kantz, H. 1994. Quantifying the Closeness of Fractal Measures, *Phys. Rev. E* 49, 5091.

Kaufman, L. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis* (Wiley, New York).

Kirkpatrick, S., C.D. Gelatt Jr., and M.P. Vecchi. 1983. Optimization by Simulated Annealing. *Science* 220, N. 4598, 671.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling. ND. *Numerical Recipes, The Art of Scientific Computing* (Book series: Cambridge Univ. Press).

Schreiber, T. 1997. Detecting and Analyzing Nonstationarity in a Time Series Using Nonlinear Cross Predictions, *Phys. Rev. Lett.* 78, 843.

Schreiber, T. and A. Schmitz. 1997. Classification of Time Series Data with Nonlinear Similarity Measures, *Phys. Rev. Lett.* 79, 1475.

Human-Machine Interaction: Potential Impact of Nanotechology in The Design of Neuroprosthetic Devices Aimed at Restoring or Augmenting Human Performance

Miguel A. L. Nicolelis, Duke University Medical Center and Mandayam A. Srinivasan, MIT

Throughout history, the introduction of new technologies has significantly impacted human life in many different ways. Until now, however, each new artificial device or tool designed to enhance human motor, sensory, or cognitive capabilities has relied on explicit human motor behaviors (e.g., hand, finger, foot movements), often augmented by automation, in order to translate the subject's intent into concrete goals or final products. The increasing use of computers in our daily lives provides a clear example of such a trend. In less than three decades, digital computers have permeated almost every aspect of our daily routine and, as a result, have considerably increased human capabilities. Yet, realization of the full potential of the "digital revolution" has been hindered by its reliance on low-bandwidth and relatively slow user-machine interfaces (e.g., keyboard, mice, etc.). Indeed, because these user-machine interfaces are far removed from the way one's brain normally interacts with the surrounding environment, the classical Von Neuman design of digital computers is destined to be perceived by the operator just as another external tool, one that needs to be manipulated as an independent extension of one's body in order to achieve the desired goal. In other words, the reach of such a tool is limited by its inherent inability to be assimilated by the brain's multiple internal representations as a continuous extension of our body appendices or sensory organs. This is a significant point, because in theory, if such devices could be incorporated into "neural space" as extensions of our muscles or senses, they could lead to unprecedented (and currently unattainable) augmentation in human sensory, motor, and cognitive performance

It is clear that recent advances in nanotechnology could significantly impact the development of brain-machine interfaces and neuroprosthetic devices. By establishing direct links between neuronal tissue and machines, these devices could significantly enhance our ability to use voluntary neuronal activity to directly control mechanical, electronic, and even virtual objects as if they were extensions of our own bodies.

Main Goals

For the past few years, we and others have proposed that a new generation of tools can be developed in the next few decades in which direct brain-machine interfaces (BMIs) will be used to allow subjects to interact seamlessly with a variety of actuators and sensory devices through the expression of their voluntary brain activity. In fact, recent animal research on BMIs has supported the contention that we are at the brink of a technological revolution, where artificial devices may be "integrated" in the multiple sensory, motor, and cognitive representations that exist in the primate brain. Such a demonstration would lead to the introduction of a new generation of actuators/sensors that can be manipulated and controlled through direct brain processes in virtually the same way that we see, walk, or grab an object.

At the core of this new technology is our growing ability to use electrophysiological methods to extract information about intentional brain processes (e.g., moving an arm) from the raw electrical activity of large populations of single neurons, and then translate these neural signals into models that control external devices. Moreover, by providing ways to deliver sensory (e.g., visual, tactile, auditory, etc.) feedback from these devices to the brain, it would be possible to establish a reciprocal (and more biologically plausible) interaction between large neural circuits and machines and hence fulfill the requirements for artificial actuators of significantly augmenting human motor performance to be recognized as simple extensions of our bodies. Using this premise and taking advantage of recent developments in the field of nanotechnology, one can envision the construction of a set of closed-loop control BMIs capable of restoring or augmenting motor performance in macro, micron, and even nano environments (Fig. C.14).

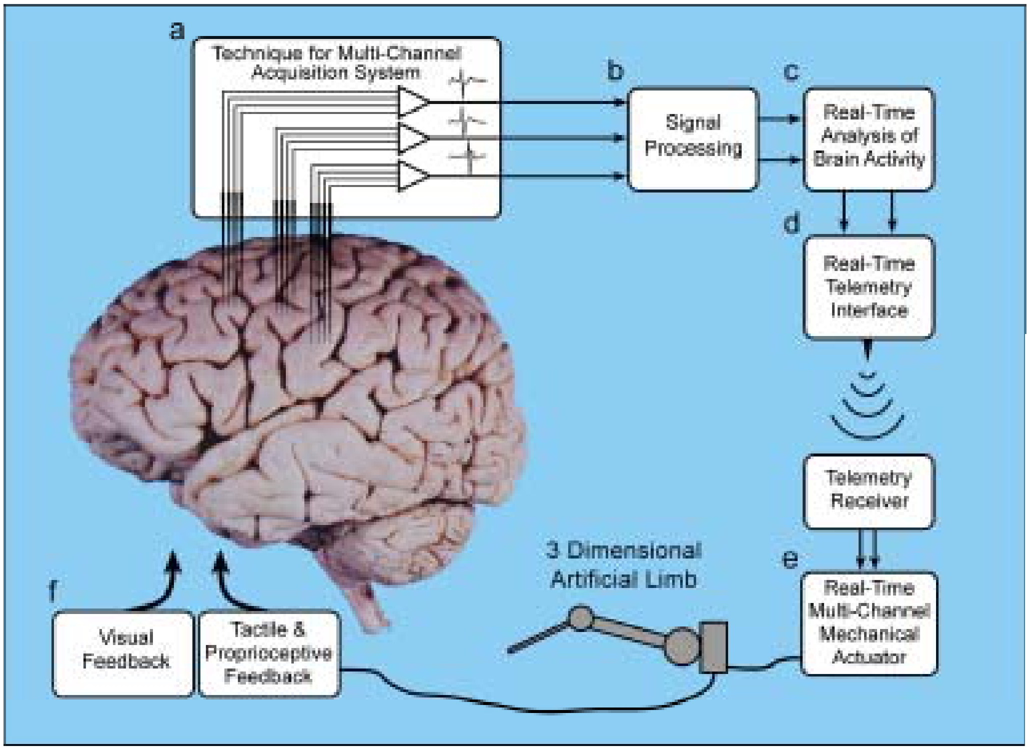


Figure C.14. General architecture of a closed-loop control brain-machine interface: Neuroprosthesis for restoring motor function of damaged brain areas.

Envisioned Utility of BMIs

The full extent to which BMIs would impact human behavior is vastly unknown. Yet, short-term possibilities are innumerable. For example, there is a growing consensus that BMIs could provide the only viable short-term therapeutic alternative to restore motor functions in patients suffering from extensive body paralysis (including lack of communication skills) resulting from devastating neurological disorders.

Assuming that noninvasive techniques to extract large-scale brain activity with enough spatial and temporal resolution can be implemented, BMIs could also lead to a major paradigm shift in the way normal healthy subjects can interact with their environment. Indeed, one can envision a series of applications that may lead to unprecedented ability to augment perception and performance in almost all human activities. These applications would involve interactions with either real or virtual environments. According to this view, real environments can also include local or remote control relative to the human subject, while virtual environments can be realistic or intentionally unrealistic. Here are some of examples.

1. **Local, real environment:** Restoration of the motor function in a quadriplegic patient. Using a neurochip implanted in the subject's brain, neural signals from healthy motor brain areas can be used to control an exoskeletal or prosthetic robotic arm used to restore fundamental motor functions such as reaching, grabbing, and walking.

2. **Remote, real environment:** Superhuman performance, such as clearing heavy debris by a robot controlled by the brain signals of a human operator located far away from the danger zone. Recent results by the P.I. and his collaborators have demonstrated that such remote control could be achieved even across the Internet.

3. **Realistic virtual environment:** Training to learn a complex sequence of repair operations by the trainee’s brain directly interacting with a virtual reality program, with or without the involvement of the trainee’s peripheral sensorimotor system.

4. **Unrealistic virtual environment:** Experiencing unrealistic physics through a virtual reality system for a “what if” scenario, in order to understand deeply the consequences of terrestrial physics.

Given the significant degree of plasticity documented even in the adult brain, repeated use of BMIs will likely transform the brain itself, perhaps more rapidly and extensively than what is currently possible with traditional forms of learning. For example, if a robot located locally or remotely is repeatedly activated via a BMI, it is likely that cortical areas specifically devoted to representing the robot will emerge, causing the robot to effectively become an extra limb of the user.

What real advantages might we obtain from future BMI based devices, compared to more conventional interfaces such as joysticks, mice, keyboards, voice recognition systems, and so forth? Three possible application domains emerge:

1. *Scaling of position and motion,* so that a “slave” actuator, being controlled directly by the subject’s voluntary brain activity, can operate within workspaces that are either far smaller (e.g., nanoscale) or far bigger (e.g., space robots; industrial robots, cranes, etc.) than our normal reach

2. *Scaling of forces and power,* so that extremely delicate (e.g., microsurgery) or high-force tasks (e.g., lifting and displacing a tank) can be accomplished

3. *Scaling of time,* so that tasks can be accomplished much more rapidly than normal human reaction time, and normally impossible tasks become possible (e.g., braking a vehicle to a stop after seeing brake lights ahead; catching a fly in your hand; catching something you have dropped; responding in hand-to-hand combat at a rate far exceeding that of an opponent)

To some extent, all these tasks, with the exception of time scaling, can, in principle, be accomplished though conventional teleoperator systems in which the human using his limbs operates a master device, which, in turn, controls a local or remote slave device. There is a history of five decades of research in this area of robotics, with moderate success, such as recent commercial development of teleoperated surgical systems. Major difficulties have been the design of appropriate master devices that the human can interact with naturally and the destabilizing effects of long time delay between the master and the slave. BMIs offer unique advantages in two ways:

1. They eliminate the need for master devices that interact with the human

2. Since the human is directly operating through his brain, the time delays associated with the signal transmission from the peripheral sensors to the CNS (~ 10–30 msec) and from CNS to the muscles (~10-30 msec), and then the time required of a limb to complete the needed action (~100-900 msec), can be reduced by an order of magnitude.

Elimination of the need for a master device is a radical departure from conventional teleoperation. Furthermore, the reduction of time delays leads to the exciting possibility of superhuman performance. For example, moving an arm from point A to point B can take ~500 msec from the time muscles are commanded by the brain, because of the force generation limitations of the muscles, the inertia of the arm, and the need to accelerate from A and to decelerate to B. But if a slave robot that is much better than the human arm in terms of power/mass ratio is directly controlled though a BMI, all three types of time delays (peripheral sensory, motor signal transmission, and limb motion) can be minimized or eliminated, possibly leading to faster and more stable operation of the slave robot. For instance, it is possible for an impaired or unimpaired person to wear an arm exoskeleton that directly interacts with the brain much faster than the natural arms.

In recent years, work developed by our laboratories has demonstrated the feasibility of building BMIs dedicated to the task of utilizing brain-derived signals to control the 1-D and 3-D movements of artificial devices. In a series of studies, we have provided the first demonstrations in animals that such BMIs can be built, that animals can learn to operate these devices in order to obtain a reward, and that motor control signals derived from the extracellular activity of relatively small populations of cortical neurons (50-100 cells) can be used to reproduce complex 3-D arm movements in a robotic device in real time.

Recent advances in nanotechnology could help significantly the advance of this area of research. First, this technology could provide new ways to extract large-scale brain activity by reducing the degree of invasiveness of current electrophysiological methods. Investment on research aimed at designing a new generation of VLSI aimed at both conditioning and analyzing large-scale electrical brain activity will also be required. Finally, a complete new generation of actuators, designed to operate in micro- or nanospaces needs to be built, since there are many new applications that can be envisioned if brain-derived signals can be employed to directly control nanomachines.

NANOTECHNOLOGY: THE MERGING OF DIAGNOSTICS AND TREATMENT

Abraham Phillip Lee, University of California at Irvine

The key to advancing from the discovery stage of nanoscience to commercially feasible nanotechnology is the ability to reliably manufacture nanoscale features and control nanoscale functions. The application of nanotechnology towards biology further requires the functional nano-interface between artificial and biological components. From a systems perspective, this requires signal transduction at matching impedances so that sensitivity and specificity are adequate to decipher the biological events. The maturation of these capabilities will enable the probing and manipulating of the fundamental building blocks of biology, namely biomolecules such as carbohydrates, lipids, nucleic acids, and proteins.

The biological cell has proven to be the most intricate functional system of its scale. Unique functionalities include its ability to regulate and adapt, hierarchical self-assembly, repair and maintainance, parallel processing, just-in-time processes, asynchronous control and signaling, and scalability from nano to macro. However, these features and functions are hard to quantify, model, engineer, and reprogram. On the other hand, microfabrication and nanofabrication techniques have given us integrated nanoscale electronics, microfluidics, microelectromechanical systems (MEMS), and microphotonics. These top-down fabrication techniques allow addressability of large-scale component platforms. On the other hand, bottom-up nanofabrication techniques (such as self-assembly) mimic how biology builds very complex systems out of simple molecules. As the scale of these two fields overlaps, devices can be developed with high sensitivity and selectivity for detecting and interfacing to biomolecules.

Projects exemplifying the field of nanobiotechnology include single molecule detection studies, functional imaging of cells and biomolecules by scanning probe microscopy, nanoparticles for targeted therapy, nanomechnical devices to measure biomolecular force interactions, etc. These research efforts represent the start towards interfacing with biological functions at the most fundamental level. However, biology is the intertwined combination of many single molecular events, each being coupled with one another either synchronously or asynchronously. To truly unveil biological events such as cell signaling pathways, genetic mutation processes, or the immune responses to pathogens, one must have a method to generate large-scale, multifunctional nano-bio interfaces with readout and control at the single biomolecule level.

I provide three visions for features of the nanobiotechnology roadmap:

1. The development of a “biological microprocessor” for synthesizing and analyzing biomolecules on nano platforms (liposomes, nanoparticles, self-assembled monolayers, and membranes) in fluids. These “biomolecular nanotransducers” will be able to function (1) as multiplexed nanomedicines capable of long duration, *in vivo* targeted detection, diagnosis, and treatment of molecular diseases; (2) as key ingredients of smart coatings for versatile environmental monitoring of toxins/pathogens; and (3) as engineered biomolecular nanosystems that mimic cellular functions for fundamental biology experiments.

2. The coupling of biomolecular units — whether they be DNA, receptors, antibodies, or enzymes — with MEMS for reassembly of cell components and reprogrammed cell functions. This will enable the rewiring of biological cell pathways in artificially controlled platforms such that it will be possible to carry out preclinical experiments without the use of animals or humans.

3. The coupling of “nano guards for health” (e.g., nanoparticles) with microfluidic controllers for long-term control of certain health parameters. For instance, the feedback loop of a glucose sensor and delivery of nano artificial islets can enable the merging of detection, diagnosis, and treatment into one MEMS device.

ARTIFICIAL BRAINS AND NATURAL INTELLIGENCE

Larry Cauller and Andy Penz, University of Texas at Dallas

It is widely accepted that nanotechnology will help push Moore’s Law to, or past, its prediction that the next few decades will witness a truly amazing advance in affordable personal computing power. Several visionary techno-futurists have attempted to estimate the equivalent power of the human brain to predict when our handheld personal computers may be able to convince us that they occasionally feel, well, unappreciated, at least. With the advent of nano-neuro-techniques, neuroscience is also about to gain unfathomable insight into the dynamical mechanisms of higher brain functions. But many neuroscientists who have dared to map the future path to an artificial brain with human intelligence do not see this problem in simple terms of “computing power” or calculations per second. We agree that the near future of nano-neuro-technology will open paths to the development of artificial brains with natural intelligence. But we see this future more in terms of a coming nano-neuro-cogno-symbiosis that will enhance human potential in two fundamental ways: (1) by creating brilliant, autonomous artificial partners to join us in our struggle to improve our world; and (2) by opening direct channels of natural communication between human and artificial nervous systems for the seamless fusion of technology and mind.

Human brain function emerges from a complex network of many billion cooperating neurons whose activity is generated by nanoscale circuit elements. In other words, the brain is a massively parallel nanocomputer. And, for the first time, nanotechnology reveals approaches toward the design and construction of computational systems based more precisely upon the natural principles of nervous systems. These natural principles include: (1) enormous numbers of elementary nonlinear computational components; (2) extensive and interwoven networks of modifiable connectivity patterns; (3) neurointeractive sensory/motor behavior; and (4) a long period of nurtured development (real or virtual). We believe human-like functions will likewise emerge from artificial brains based upon these natural principles.

A simple nanoelectronic component, the resonant tunneling diode, possesses nonlinear characteristics similar to the channel proteins that are responsible for much of our neurons' complex behavior. In many ways, nanoscale electronics may be more suitable for the design of nonlinear neural networks than as simple switching elements in digital circuits. At this NBIC meeting, Phil Kuekes from Hewlett-Packard described a nanoscale cross-link connection scheme that may provide an approach to solving the truly difficult problem of how to interconnect enormous networks of these nanocomponents. But as a beginning, these initial steps to realization of a nano-neuro-computer permit a consideration of the much greater density that is possible using nanoelectronic neurons than has so far been possible with microelectronic solutions, where equivalent chip architectures would need to be millions of times larger. If the size of the artificial brain were small enough to mount on a human-size organism, then it may be simpler to design nurturing environments to promote the emergence of human-like higher functions.

Decades of neuroscience progress have shed a great deal of light upon the complexity of our brain's functional neuro-architecture (e.g., Felleman and Van Essen 1991). Despite its extreme complexity (>100,000 miles of neuron fibers), fundamental principles of organization have been established that permit a comprehensive, although highly simplified sketch of the structure responsible for natural intelligence. In addition, neuroscience has characterized many of the principles by which the network's connections are constantly changing and self-organizing throughout a lifetime of experience (e.g., Abbott and Nelson 2001). While some futurists have included the possibility that it will be possible to exactly replicate the cellular structure of the human brain (Kurzweil 1999), it seems impossible from a neuroscience point of view, even with nanotechnology. But it is not necessary to be too precise. Genetics is not that precise. We know many of the principles of neuro-competition and plasticity that are the basis for the continuous refinement of neural functions in the midst of precise wiring and environmental complexity. But the only test of these far-reaching principles is to construct a working model and learn to use it.

Constrained by the limits of microtechnology, previous attempts to mimic human brain functions have dealt with the brain's extreme complexity using mathematical simplifications (i.e., neural networks) or by careful analysis of intelligent behavior (i.e., artificial intelligence). By opening doors to the design and construction of realistic brain-scale architectures, nanotechnology is allowing us to rethink approaches to human-like brain function without eliminating the very complexity that makes it possible in the first place. The tools of nonlinear dynamical mechanics provide the most suitable framework to describe and manage this extreme complexity (e.g., Kelso 1995; Freeman 2000). But the first step is to recognize and accept the natural reality that the collective dynamics of the neural process responsible for the highest human functions are not mathematically tractable.

Instead, higher functions of the brain are emergent properties of its neuro-interactivity between neurons, between collections of neurons, and between the brain and the environment. While purely deterministic, it is no more possible to track the cause-effect path from neuron activity to higher functions such as language and discovery than it is to track the path from an H_2O molecule to the curl of a beach wave. Unfortunately, appeals to emergence always leave an unsatisfying gap in any attempt to provide a complete explanation, but nature is full of examples, and classical descriptions of human intelligence have depended strongly upon the concept of emergence (i.e., Jean Piaget, see Elman et al. 1997). But modern emergent doctrine is gaining legitimacy from the powerful new tools of nonlinear dynamical mathematics for the analysis of fractals and deterministic chaos. Instead of tracking cause-effect sequence, the new paradigm helps to identify dynamical mechanisms responsible for the phase shifts from water to ice, or from exploring to understanding.

From the perspective of neuro-interactive emergence, brain function is entirely self-organized so it may only be interpreted with respect to the interactive behavior of the organism within meaningful contexts. For instance, speech communication develops by first listening to one's own speech sounds, learning to predict the sensory consequence of vocalization, and then extending those predictions to include the response of other speakers to one's own speech. This natural process of self-growth is radically different from the approaches taken by artificial intelligence and "neural net" technologies. The kernel of this natural process is a proactive hypothesis-testing cycle spanning the scales of the nervous system that acts first and learns to predict the resulting consequences of each action within its context (Cauller, in press; see also Edelman and Tonomi 2001). Higher functions of children emerge as a result of mentored development within nurturing environments. And emergence of higher functions in artificial brains will probably require the same kinds of care and nurturing infrastructure we must give our children.

So the future of the most extreme forms of machine intelligence from this neuroscience perspective differs in many respects from popular visions: (1) "artificial people" will be very human-like given their natural intelligence will develop within the human environment over a long course of close relationships with humans; (2) artificial people will not be like computers any more than humans are. In other words, they will not be programmable or especially good at computing; (3) artificial people will need social systems to develop their ethics and aesthetics.

An optimal solution to the problem of creating a seamless fusion of brain and machine also needs to be based upon these neurointeractive principles. Again, nanotechnology, such as minimally invasive nano-neuro transceivers, is providing potential solutions to bridge the communication gap between brain and machine. But the nature of that communication should be based upon the same neural fundamentals that would go into the design of an artificial brain.

For instance, sensory systems cannot be enhanced by simply mapping inputs into the brain (e.g., stimulating the visual cortex with outputs from an infrared camera won't work). The system must be fused with the reciprocating neurointeractivity that is responsible for ongoing conscious awareness. This means that brain control over the sensory input device is essential for the system to interpret the input in the form of natural awareness (e.g., there must be direct brain control over the position of the video source). In other words, brain enhancements will involve the externalization of the neurointeractive process into peripheral systems that will respond directly to brain signals. These systems will become an extension of the human mind/body over a course of accommodation that resembles the struggle of physical therapy following cerebral stroke.

Fusion of artificial brains into larger brains that share experience is a direct extension of this line of reasoning. This also would not be an immediate effect of interconnection, and the fusion would involve give and take on both sides of the connection over an extended course of active accommodation. But the result should surpass the sum of its parts with respect to its ability to cope with increasing environmental complexity.

Speculation leads to the next level of interconnection, between human and artificial brains. On the face of it, this appears to be a potential path to cognitive enhancement. However, the give and take that makes neurointeractive processes work may be too risky when humans are asked to participate.

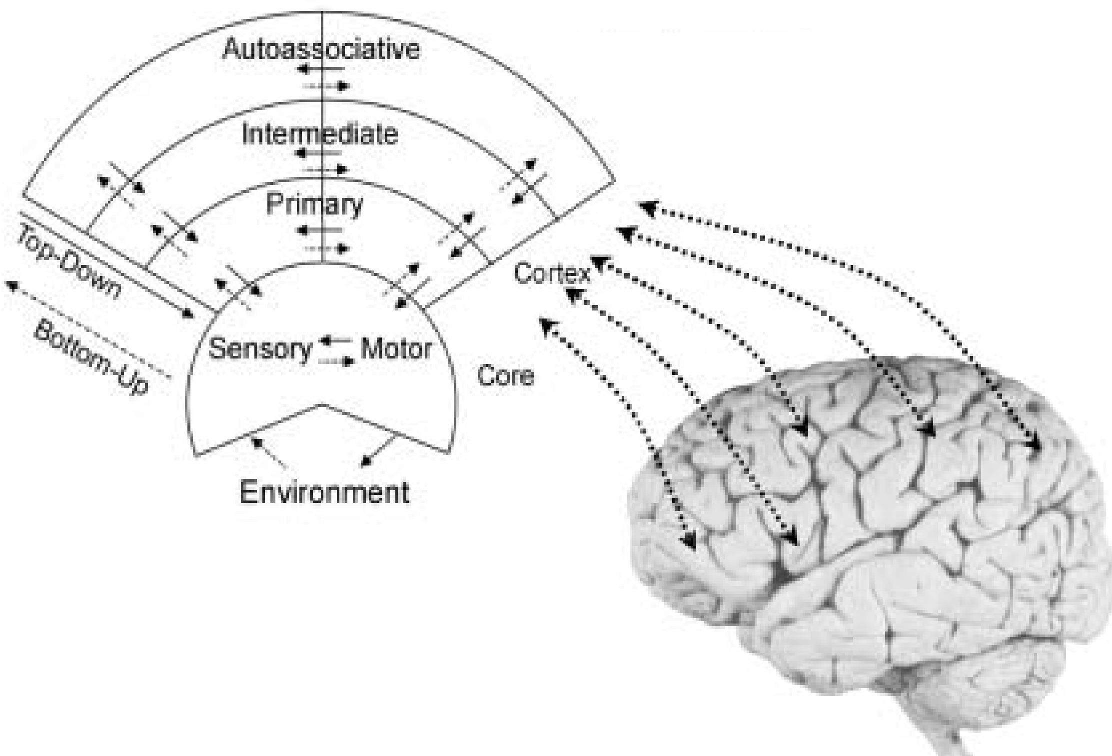


Figure C.15. Neurointeractive artificial brain/human brain interface for neuroprosthesis or enhancement.

References

Abbott, L.F., and Nelson SB. 2000. Synaptic plasticity: taming the beast. *Nat Neurosci 3:*1178-83

Cauller, L.J. (*in press*). The neurointeractive paradigm: dynamical mechanics and the emergence of higher cortical function. In: *Theories of Cerebral Cortex,* Hecht-Neilsen R and McKenna T (eds).

Edelman G.M. and G. Tonomi. 2001. *A Universe of Consciousness: How Matter Becomes Imagination,* Basic Books.

Elman, J.L., D. Parisi, E.A. Bates, M.H. Johnson, A. Karmiloff-Smith. 1997. *Rethinking Innateness: A Connectionist Perspective on Development*, MIT Press, Boston.

Felleman, D.J., and Van Essen D.C. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex 1(1):*1-47.

Freeman, W.J. 2000. *Neurodynamics: An Exploration in Mesoscopic Brain Dynamics (Perspectives in Neural Computing).* Springer Verlag.

Kelso, S. 1995. *Dynamic Patterns (Complex Adaptive Systems)*. MIT Press, Boston, MA.

Kurzweil, R. 1999. *The Age of Spiritual Machines*. Viking Press, New York, NY.

CONVERGING TECHNOLOGIES FOR PHYSIOLOGICAL SELF-REGULATION

Alan T. Pope, NASA Langley Research Center, and Olafur S. Palsson, Mindspire, LLC

The biofeedback training method is an effective health-enhancement technique, which exemplifies the integration of biotechnology and information technology with the reinforcement principles of cognitive science. Adding nanotechnology to this mix will enable researchers to explore the extent to which physiological self-regulation can be made more specific and even molecular, and it may lead to a entire new class of effective health-enhancing and health-optimizing technologies.

Vision

Physiological Self-Regulation Training

Biofeedback is a well-established and scientifically validated method to treat a variety of health problems and normalize or enhance human physiological functioning. It consists of placing sensors on the body to measure biological activity, and enabling patients to self-correct their physiological activity by showing them on a computer screen (typically in the form of dynamic graphs) what is going on inside their bodies.

> Biofeedback means "the feeding back of information to the individual about change in a physiological system." It implies that the subject is continuously, or discontinuously, informed about change in a particular physiological system under study. The information is believed to act as a reinforcer for further changes in either the same or the opposite direction. As a result of instrumental learning, a physiological response may come under "instructional" or "volitional" control as a function of the feedback of information. (Hugdahl 1995, 39)

When patients are able to observe the moment-to-moment changes in their physiological activity in this way, they can learn over time to control various body functions that are usually outside of conscious control, such as heart rate, muscle tension, or blood flow in the skin:

> According to a basic premise in biofeedback applications, if an individual is given information about biological processes, and changes in their level, then the person can learn to regulate this activity. Therefore, with appropriate conditioning and training techniques, an individual can presumably learn to control body processes that were long considered to be automatic and not subject to voluntary regulation. (Andreassi 2000, 365)

Biofeedback has been used for forty years with considerable success in the treatment of various health problems, such as migraine headaches, hypertension, and muscle aches and pains. More recently, biofeedback training has been used to enhance performance in a number of occupations and sports activities (Norris and Currieri 1999). At NASA Langley Research Center, work in physiological self-regulation is directed at reducing human error in aviation:

> Our work has focused on a number of areas with the goal of improving cognitive resource management, including that of physiological self-regulation reported here. Other areas include adaptive task allocation, adaptive interfaces, hazardous unawareness modeling, cognitive awareness training, and stress-counter-response training. (Prinzel, Pope, and Freeman 2002, p. 196)

Intrasomatic Biofeedback: A New Frontier

The exclusive reliance upon sensing of physiological functions from the surface of the body has limited biofeedback's specificity in targeting the physiological processes that underlie human performance and the physiological dysregulation implicated in several disorders. Biofeedback technology has yet to incorporate recent advances in biotechnology, including nanoscale biosensors, perhaps because biofeedback research and practice is dominated by a focus on traditional and proven training protocols rather than on biotechnology.

> As a result of the development of new analytical tools capable of probing the world of the nanometer, it is becoming increasingly possible to characterize the chemical and mechanical properties of cells (including processes such as cell division and locomotion) and to measure properties of single molecules. These capabilities complement (and largely supplant) the ensemble average techniques presently used in the life sciences. (Roco and Bainbridge 2001, 7)

Current biofeedback technology still mostly detects, processes, and feeds back to trainees broad signals from sensors on the skin. Such surface sensors are only suited for providing summary information about broad functional characteristics of the organism, like overall autonomic functioning, summative brain activity in a large portion of the cortex, or activity levels of large masses of striated muscle.

Nanoscale technologies, specifically nanoscale biosensor technology, hold the potential for realtime sensing and feedback of internal bodily processes that are the origins or precursors of the physiological signals sensed on the skin surface by current biofeedback technology. Intrasomatic signals, closer to the physiological source of the body activity of interest than surface-detectable signals, could be used for more targeted and precise feedback conditioning of physiological functions and physiological dysregulation. They could also be used to dynamically feed back to patients the consequences and benefits of exercises and practices, or warnings of hazardous alterations in physiology, in order to provide education as well as motivation for adhering to prescribed behavioral treatment regimens. Furthermore, the presence of such small intrasomatic sensors could enable physicians or surveillance computers to titrate or fine-tune the treatment of a patient's disorder (such as medication flow-rate) in ways otherwise not possible.

Early work by Hefferline, Keenan, and Harford (1959) demonstrated that covert physiological responses could be conditioned by attaching consequences, in a traditional psychological reinforcement paradigm, to the production of the responses without the trainee's conscious, deliberate effort to control the responses. Most biofeedback training successes do indeed operate without the necessity for the trainee to be able to articulate the exact nature of the efforts they employ in the learning process, and sometimes without them even trying to consciously control the process. Nevertheless, an additional application of feedback of nanoscale biosensed parameters may be to inform the trainee of the results of his/her overt efforts to facilitate management of a physiological function. An example would be the moment-to-moment feedback of blood oxygenation level or oxygen/CO_2 balance in respiration training for hyperventilation in panic disorder (Ley 1987).

Roles of Converging Technologies

The roles of NBIC technologies in the intrasomatic biofeedback vision are illustrated schematically in Figure C.16.

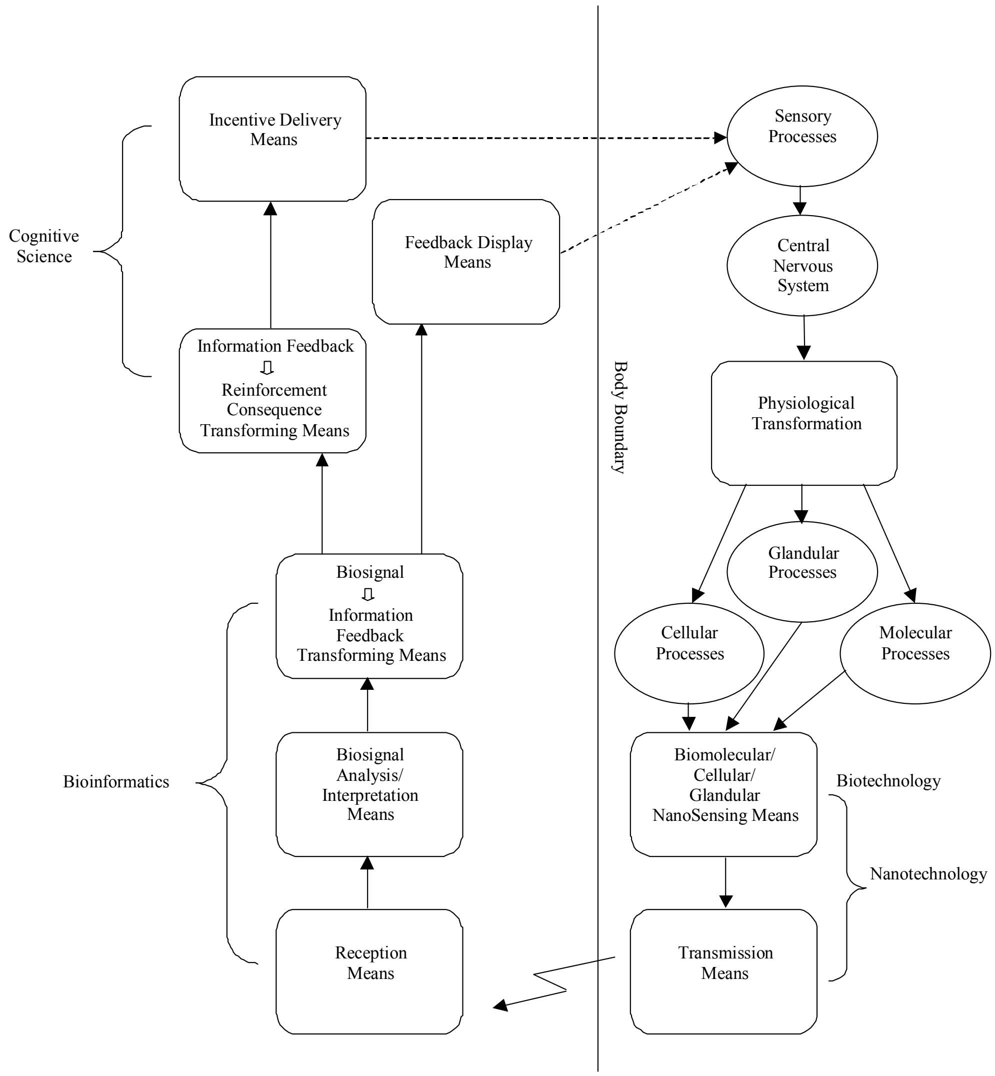


Figure C.16. Intrasomatic biofeedback.

Cognitive Science

Mainly used in psychophysiology as an applied technique, the principle of biofeedback goes back to the idea that nonvolitional, autonomic behavior can be instrumentally conditioned in a stimulus-reinforcement paradigm.

Traditional learning theory at the time of the discovery of the biofeedback principle held that an autonomic, involuntary response could be conditioned only through the principles of classical, or Pavlovian, conditioning. Instrumental, operant learning could be applied only to voluntary behavior and responses. However, in a series of experiments, Miller (1969) showed that autonomic behavior, like changes in blood pressure, could be operantly conditioned in rats (Hugdahl 1995, 40).

In the beginning of the biofeedback field, researchers, working with animals, experimented with more precisely accessing internal physiological phenomena to provide the signals and information representing the functions to be conditioned:

> The experimental work on animals has developed a powerful technique for using instrumental learning to modify glandular and visceral responses. The improved training technique consists of moment-to-moment recording of the visceral function and immediate reward, at first, of very small changes in the desired direction and then of progressively larger ones. The success of this technique suggests that it should be able to produce therapeutic changes (Miller 1969, 443-444).

Miller identified critical characteristics that make a symptom (or physiological function) amenable to instrumental conditioning through biofeedback:

> Such a procedure should be well worth trying on any symptom, functional or organic, that is under neural control, that can be continuously monitored by modern instrumentation, and for which a given direction of change is clearly indicated medically — for example, cardiac arrhythmias, spastic colitis, asthma, and those cases of high blood pressure that are not essential compensation for kidney damage (Miller 1969, 443-444).

The mechanism of neural control that would enable instrumental conditioning of basic molecular physiological processes has yet to be identified. Current understanding is limited to the notion that it generally involves a "bucket brigade" effect where willful cognitive influences in the cortex are handed down through the limbic system and on down into the hypothalamus, which disseminates the effect throughout the body via various neural and endocrine avenues.

Similarly, researchers in the field of psychoneuroimmunology have yet to find the exact biological mechanisms linking the brain and the immune system. Nevertheless, Robert Ader, one of the first to present evidence that immune responses could be modified by classical conditioning (Ader and Cohen 1975), states:

> There are many psychological phenomena, and medical phenomena for that matter, for which we have not yet defined the precise mechanisms. It doesn't mean it's not a real phenomenon (Azar 1999).

Nanobiotechnology

Miller's (1969, 443-444) requirement that the physiological function be "continuously monitored by modern instrumentation" is now made possible by nanoscale biosensors, enabling the investigation of the instrumental conditioning of biomolecular phenomena.

> Implantable sensors or "smart" patches will be developed that can monitor patients who are at risk for specific conditions. Such sensors might monitor, for example, blood chemistry, local electric signals, or pressures. The sensors would communicate with devices outside the body to report results, such as early signals that a tumor, heart damage, or infection is developing. Or these sensors could be incorporated into "closed loop" systems that would dispense a drug or other agent that would counteract the detected anomaly. For chronic conditions like diabetes, this would constitute a great leap forward. Nanotechnology will contribute critical technologies needed to make possible the development of these sensors and dispensers (NSTC 2000, 54, 55).

Another "closed loop system" that would "counteract the detected anomaly" is intrasomatic biofeedback training. In this case, remediation of a physiological anomaly or suboptimal condition would be achieved by self-regulation learned through instrumental conditioning, rather than by an external agent such as a drug or nanodevice.

Freitas (1999, section 4.1) describes "nanosensors that allow for medical nanodevices to monitor environmental states at three different operational levels," including "local and global somatic states (inside the human body)," and cellular bioscanning:

> The goal of cellular bioscanning is the noninvasive and non-destructive in vivo examination of interior biological structures. One of the most common nanomedical sensor tasks is the scanning of cellular and subcellular structures. Such tasks may include localization and examination of cytoplasmic and nuclear membranes, as well as the identification and diagnostic measurement of cellular contents including organelles and other natural molecular devices, cytoskeletal structures, biochemical composition and the kinetics of the cytoplasm (Freitas 1999, section 4.8).

The function of "communicating outside the body to report results" (NSTC 2000, 54, 55) is essential for an intrasomatic biofeedback application. Freitas (1999) describes a similar function for nanorobots:

> In many applications, in vivo medical nanodevices may need to communicate information directly to the user or patient. This capability is crucial in providing feedback to establish stable and reliable autogenous command and control systems (Chapter 12). Outmessaging from nanorobot to the patient or user requires the nanodevice to manipulate a sensory channel that is consciously available to human perception, which manipulation can then be properly interpreted by the patient as a message.
>
> Sensory channels available for such communication include sight, audition, gustation and olfaction, kinesthesia, and somesthetic sensory channels such as pressure, pain, and temperature (Freitas 1999, section 7.4.6).

In this application, "outmessaging" is described as enabling user control of a nanorobot; for intrasomatic biofeedback, this function would provide the information that acts as a reinforcer for conditioning changes in cellular and molecular processes (Figure C.16).

Transforming Strategy

A Technical Challenge

Early on, Kamiya (1971) specified the requirements for the biofeedback training technique, and these have not changed substantially:

- The targeted physiological function must be monitored in real time.
- Information about the function must be presented to the trainee so that the trainee perceives changes in the parameter immediately.
- The feedback information should also serve to motivate the trainee to attend to the training task.

The challenges for the fields of nanotechnology, biotechnology, information technology, and cognitive science (NBIC) in creating the technology to enable internally targeted physiological self-regulation technology can be differentiated according to the disparities between (1) the time response of existing physiometric technology, (2) the time course of the targeted physiological processes, and (3) the requirements for feedback immediacy in the biofeedback paradigm. Realtime sensing is essential to make the processes available for display and attaching sensory feedback consequences to detected changes.

The physiological processes most readily amenable to biofeedback self-regulation are those where the internal training targets are available in real time with current or emerging technologies, such as electrical (e.g., brainwave) and hydraulic (e.g., blood flow) physiological signals.

Instruments using microdialysis, microflow, and biosensor technologies to deliver blood chemistry data such as glucose and lactate in real time (European Commission 2001) will need to reduce test cycle time from minutes to seconds to meet the feedback immediacy criterion required for biofeedback training. Even then, it may be discovered that time delays between the initiation of the production of these chemicals and their appearance in the bloodstream require that signals from upstream stages in the formation process are more appropriate targets for feedback in the self-regulation training loop.

Flow cytometry is an example of an offline, non-realtime technology, in this case for measuring certain physical and chemical characteristics, such as size, shape, and internal complexity, of cells or particles as they travel in suspension one by one past a sensing point. For the blood cell formation process that controls these characteristics of cells, hematopoiesis, to become a candidate for physiological self-regulation training will require advances in molecular-scale technology. These advances will probably need to occur in the upstream monitoring of molecular or biosignal (hormonal, antibody, etc.) precursors of the blood cell formation process, bringing tracking of the process into the realtime scale required for feedback immediacy.

Internal nanosensors will similarly solve the time-response problem that has prevented the utilization of brain functional monitoring and imaging in biofeedback.

> Thus, current functional imaging methods are not in real time with brain activity; they are too slow by a factor of 100 or more. The big advance will be to develop functional imaging techniques that show us — as it is happening — how various areas of the brain interact. … Do not ask me what the basis of this new imaging will be. A combination of electrical recording and changes in some other brain properties perhaps? (McKhann 2001, 90)

> The precision and speed of medical nanodevices is so great that they can provide a surfeit of detailed diagnostic information well beyond that which is normally needed in classical medicine for a complete analysis of somatic status (Freitas 1999, section 4.8).

Enabling Collaborations

The collaboration of key institutions will be necessary to expedite the development of the intrasomatic biofeedback vision. Potentially enabling joint efforts are already in place (National Aeronautics and Space Administration [NASA] and the National Cancer Institute [NCI] 2002):

> NASA and the National Cancer Institute (NCI) cosponsor a new joint research program entitled Fundamental Technologies for the Development of Biomolecular Sensors. The goal of this program is to develop biomolecular sensors that will revolutionize the practice of medicine on Earth and in space.

> The Biomolecular Systems Research Program (BSRP) administrates the NASA element of the new program, while the Unconventional Innovations Program (UIP) does so for NCI.

> NASA and NCI are jointly seeking innovations in fundamental technologies that will support the development of minimally invasive biomolecular sensor systems that can measure, analyze, and manipulate molecular processes in the living body. (National Aeronautics and Space Administration [NASA] 2002)

One of the purposes that this program is designed to serve is NASA's requirement "for diagnosis and treatment of injury, illness, and emerging pathologies in astronauts during long duration space missions … Breakthrough technology is needed to move clinical care from the ground to the venue of long duration space flight … Thus, the space flight clinical care system must be autonomous …" (NASA/NCI 2001). Intrasomatic biofeedback's potential for self-remediation of physiological changes that threaten health or performance would be useful in many remote settings.

The nanotechnology, biotechnology, and information technology (NBI) components of the NASA/NCI joint project are specified in a NASA News Release:

> The ability to identify changes such as protein expression or gene expression that will develop into cancer at a later date may enable scientists to develop therapies to attack these cells before the disease spreads. "With molecular technologies, we may be able to understand the molecular signatures within a cell using the fusion of biotechnology, nanotechnology and information technology," [John] Hines [NASA Biomolecular Physics and Chemistry Program Manager] said.
>
> [NASA] Ames [Research Center] will focus on six key areas in molecular and cellular biology and associated technologies. Biomolecular sensors may some day be able to kill tumor cells or provide targeted delivery of medication. Molecular imaging may help scientists understand how genes are expressed and how they control cells. Developments in signal amplification could make monitoring and measurement of target molecules easier. Biosignatures — identification of signatures of life — offer the possibility of distinguishing cancerous cells from healthy cells. Information processing (bioinformatics) will use pattern recognition and modeling of biological behavior and processes to assess physiological conditions. Finally, molecular-based sensors and instrumentation systems will provide an invaluable aid to meeting NASA and NCI objectives (Hutchison 2001).

The NASA/NCI project is designed to "develop and study nanoscale (one-billionth of a meter) biomedical sensors that can detect changes at the cellular and molecular level and communicate irregularities to a device outside the body" (Brown 2001). This communication aspect of the technology will make possible the external sensory display of internal functioning that is essential to the intrasomatic biofeedback vision (Figure C.16).

Collaborations such as this NASA/NCI project provide the NBI components of the intrasomatic biofeedback vision. The participation of organizations devoted to the development and application of cognitive science (C), such as those specified in Figure C.17, would complete the set of disciplines necessary to realize the vision.

Estimated Implications: The Promise of Intrasomatic Biofeedback

It has not been widely appreciated outside the highly insular field of psychophysiology that humans, given sufficiently informative feedback about their own physiological processes, have both the capacity and inherent inclination to learn to regulate those processes. This phenomenon has, however, been established conclusively in numerous biofeedback applications across a range of different biological functions, including the training of brain electrical activity and of autonomic responses. The integration of NBIC technologies will enable the health- and performance-enhancing benefits of this powerful methodology to be extended to other critical physiological processes not previously considered amenable to change by training.

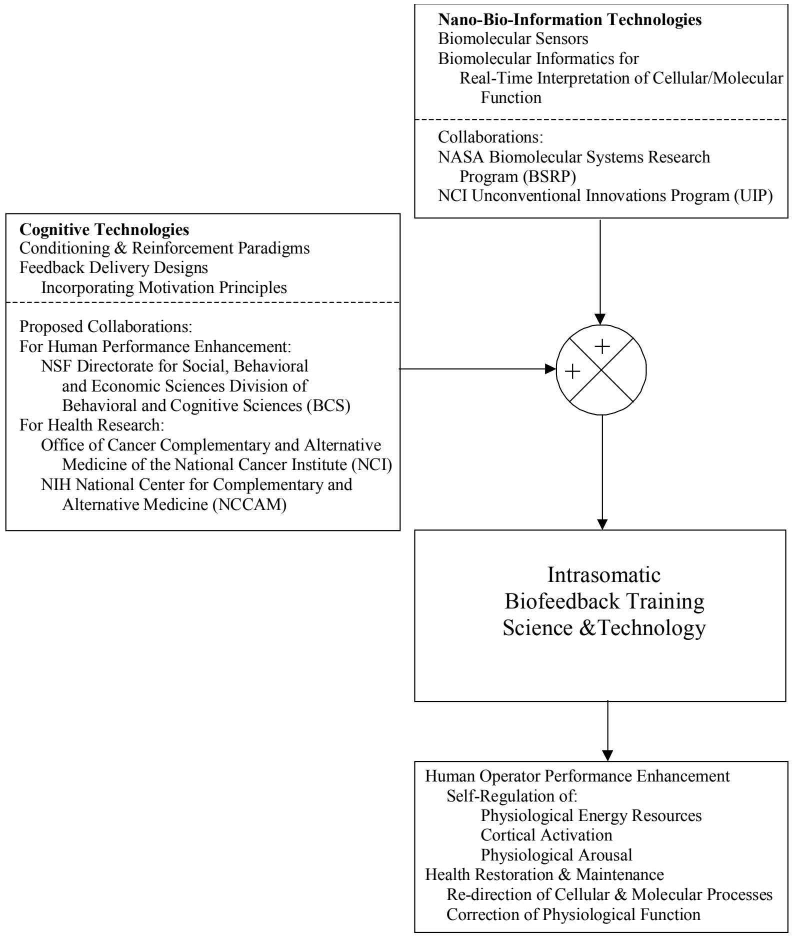


Figure C.17. Enabling collaborations.

While self-regulation of basic molecular physiological processes may seem fantastical at the present time, it is worth keeping in mind that therapeutic conditioning of autonomic and brainwave signals, now well established, was similarly considered in the fantasy realm no more than four decades ago. The discovery of the human capacity for physiological self-regulation awaited the inventiveness of pioneers, who, in a bold empowering stroke, displayed physiological signals, previously scrutinized only by the researcher, to the subjects whose signals they were, with the aim of giving the subjects control of these processes. This innovation began the discovery process that has demonstrated that, given the right information about their bodily processes in the right form, people can exert impressive control over those responses. The integration of NBIC with the biofeedback method opens an entirely new frontier, inviting the pioneers of a new era in psychophysiology to explore the extent to which this physiological self-regulation can be made more precise, perhaps even to the point of reliably modifying specific molecular events. These developments will enable human beings to willfully induce inside their own bodies small and highly specific biological changes with large health- and performance-enhancing consequences.

References

Ader R. and Cohen N. 1975. Behaviorally conditioned immunosuppression. *Psychosomatic Medicine*, 37(4): 333-340.

Andreassi, J. L. 2000. Psychophysiology: human behavior and physiological response, 4th edition, New Jersey: Lawrence Erlbaum Associates.

Azar, B. (1999). Father of PNI reflects on the field's growth. *APA Monitor*, 30(6). Retrieved April 7, 2002 from http://www.apa.org/monitor/jun99/pni.html.

Brown, D. 2001. "Joint NASA/NCI Research to Develop Sensors for Health Monitoring Inside the Human Body". News Release 01-229, Nov. 21, 2001, NASA Headquarters, Washington, DC. Retrieved May 3, 2002 from http://spaceresearch.nasa.gov/general_info/OBPR-01-230.html.

European Commission. 2001. Blood Chemistry in Real Time. Innovation in Europe: Research and Results: Medicine and Health. Retrieved November 5, 2001, from http://europa.eu.int/comm/research/success/en/med/0011e.html.

Freitas, R.A., Jr. 1999. Nanomedicine, Volume I: Basic Capabilities. Landes Bioscience. Retrieved April 7, 2002 from http://www.landesbioscience.com/nanomedicine/.

Hefferline, R.F., Keenan, B., and Harford, R.A. 1959. Escape and Avoidance Conditioning in Human Subjects without Their Observation of the Response. *Science*, 130, 1338-1339.

Hugdahl, K. 1995. *Psychophysiology: The Mind-Body Perspective*. Cambridge, MA: Harvard University Press.

Hutchison, A. 2001. "NASA Biotechnology Project May Advance Cancer Research." News Release 01-96AR, Dec. 5, 2001, NASA Ames Research Center, Moffett Field, Calif. Retrieved May 3, 2002 from http://amesnews.arc.nasa.gov/releases/2001/01_96AR.html.

Kamiya, J. 1971. Biofeedback and Self-Control: Preface. Chicago: Aldine-Atherton. ix-xvi.

Ley, R. 1987. Panic Disorder: A Hyperventilation Interpretation. In Michelson, L. and Ascher, L.M. (eds.). Anxiety and Stress Disorders: Cognitive-Behavioral Assessment and Treatment. New York: The Guilford Press. 191-212.

McKhann, G. M. 2001. A Neurologist Looks Ahead to 2025. *Cerebrum*, 3(3), 83-104.

Miller, N. E. 1969. Learning of Visceral and Glandular Responses. *Science*, 163, 434-445.

National Aeronautics and Space Administration (NASA). 2002. BioMolecular Systems Research Program. NASA AstroBionics Program. Retrieved April 7, 2002 from http://astrobionics.arc.nasa.gov/prog_bsrp.html.

National Aeronautics and Space Administration (NASA) and the National Cancer Institute (NCI). 2002. Biomolecular Sensor Development: Overview. Retrieved May 3, 2002 from http://nasa-nci.arc.nasa.gov/overview_main.cfm.

National Cancer Institute (NCI) and the National Aeronautics and Space Administration (NASA). 2001. "Fundamental Technologies for Development of Biomolecular Sensors." NASA/NCI Broad Agency Announcement (BAA) (N01-CO-17016-32). Retrieved April 7, 2002 from http://rcb.nci.nih.gov/appl/rfp/17016/Table%20of%20Contents.htm.

National Science and Technology Council Committee on Technology, Subcommittee on Nanoscale Science, Engineering and Technology. 2000. National Nanotechnology Initiative: The Initiative and its Implementation Plan. Vision: Advanced Healthcare, Therapeutics and Diagnostics: a. Earlier Detection and Treatment of Disease: Sensors. Washington, D.C.

Norris, S. L., and Currieri, M. Performance Enhancement Training Through Neurofeedback. 1999. In Evans, J. R. and Abarbanel, A. (eds.) Introduction to Quantitative EEG and Neurofeedback. San Diego: Academic Press. 223-240.

Prinzel, L. J., Pope, A. T., and Freeman, F. G. 2002. Physiological Self-Regulation and Adaptive Automation. *The International Journal of Aviation Psychology*, 12(2), 181-198.

Roco, M. C. and Bainbridge, W. S. (eds.). 2001. Societal Implications of Nanoscience and Nanotechnology. Dordrecht, Netherlands: Kluwer Academic Publishers.

IMPROVING QUALITY OF LIFE OF DISABLED PEOPLE USING CONVERGING TECHNOLOGIES

G. Wolbring, U. Calgary, and R. Golledge, UCSB

It is understood that NBIC should be used in a way that diminishes the discrimination against disabled people, advances their acceptance and integration into society, and increases their quality of life.

The Vision

1. *NBIC has the potential to give disabled people, and this includes many elderly, the ability to choose between different modes of information output*, whether visual, audio, print, or others, as all of these modes can be offered routinely at the same time. It has the potential to change computer interface architecture so that disabled people, including those who are blind, sight-impaired, dyslexic, arthritic, immobile, and deaf, can access the Internet and its webpages as transparently and quickly as able-bodied people by means of, for example, holographic outputs; force-feedback, vibrotactile, vastly improved natural speech interfaces; and realtime close captioning. Multimodal access to data and representations will provide a cognitively and perceptually richer form of interaction for all persons, regardless of impairment, handicap, or disability. It will allow for more flexibility in the mode of working (from home or a company building or elsewhere) and representation (in person or virtual). Meetings like this workshop could easily take place within a 3-D virtual reality once the modes of interaction are available in real time and adaptable to different needs (see e.g., http://www.digitalspace.com/avatars/). Even private conversations during the breaks could be easily arranged in this virtual reality. This virtual reality would be an alternative to travel. Multimodal input and output interfaces will allow human-computer (HC) interaction when sight is not available (e.g., for blind or sight-impaired users), when sight is an inappropriate medium (e.g., accessing computer information when driving a vehicle at high speeds), or when features and objects are occluded or distant.

2. *NBIC has the potential to increase the quality of life of disabled people by allowing for alternative modes of transportation*. One technique that could potentially increase quality of life immensely would be mobile teleportation devices. Teleportation would be linked to global positioning devices (see http://www.research.ibm.com/quantuminfo/teleportation/) so that someone could just teleport themselves where they have to go.

3. *NBIC will allow for improving assistive devices for disabled people.* For example, wheelchairs, which so far haven't changed much in the last 20 years, could be improved in several ways: nanomaterials could make them more durable, cheaper, and lighter; nanotechnology can be used to improve batteries or develop alternative energy generating devices (such as small fuel cells); NBIC could increase wheelchair capabilities (such as stair climbing) and make them more intelligent. The resulting device would allow a person sitting in it to move in any direction, horizontal or vertical, without regard to obstacles such as stairs. It have no need to physically attach to a surface for movement (it could hover). It would allow for the exploration of rough terrain such as the outdoors. This kind of personal moving/flying device could of course be developed for all people. NBIC also might lead to functional artificial limbs, which might even be better than existing human limbs. The same is true for the development of artificial devices for hearing, vision, and cognitive abilities such as comprehension and memory.

4. *NBIC will greatly improve the functionality and design of houses,* allowing voice command, intelligent applications, etc., that enable disabled (and elderly) people to be more independent.

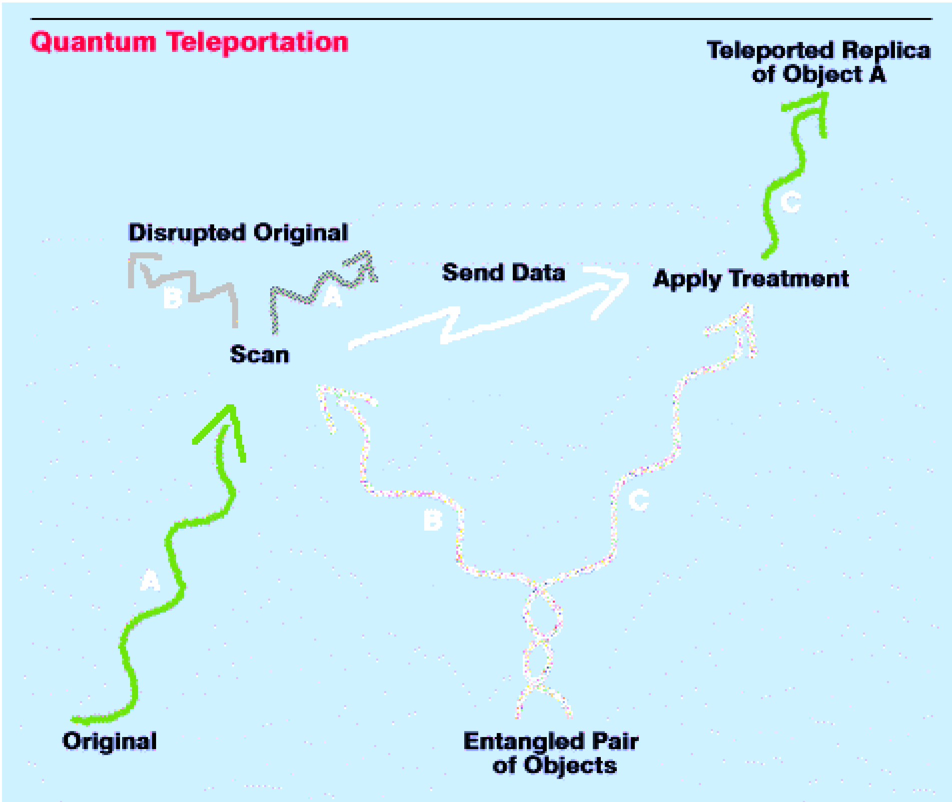


Figure C.18. On the quantum level this transport is achievable (Shahriar, Shapiro and Hemmer 2001). A mobile human teleportation device that can transport the person wherever the person wants to be would solve many accessibility and transportation problems.

5. *NBIC has the potential to change the public space to make it much more user friendly and inclusive.* Means will include IT advances to enable wearable computers for use in everyday living (e.g., finding when the next bus is due or where it is now); creation of smart environments (e.g., Remote Auditory Signage Systems [RASS] like talking signs, talking buses, etc., to facilitate wayfinding, business/object location identification, recognition of mass transit services, and intermodal transfer); use of IT and cognitive technology to develop voice-activated personal guidance systems using GPS and GIS; and multimodal interfaces to assist travel and environmental learning.

6. *NBIC has the potential to improve communication on a global scale* (e.g., universal translation devices), which would allow for a greater exchange of knowledge between people and a faster dissemination of advances in NBIC. The devices available today are not accurate and intelligent enough for use in day-to-day communication.

7. *NBIC has the potential to help in the health management of disabled — and all — people*.

The Role of Converging Technologies

The converging of technologies is needed if a systematic approach is to be undertaken to use technology for the benefit of disabled people. Often the same tool will have to rely on more than one technology to be workable (e.g., a wheelchair needs improved nanomaterials science for weight reduction and IT and cogno-science for new forms of control, leading to a whole new type of moving device such as a personal moving/flying device.)

The Transforming Strategy

The transforming strategy starts with the goal to increase the quality of life of disabled people. This goal makes it self-evident that disabled people have to be present at every brainstorming on every level, whether in government or private companies or in the public. These brainstorming activities will lead to the generation of ideas and identification of solutions for the goal. The generation of ideas and identifications leads to the identification of the technologies needed to implement these ideas and solutions. Technology is all the time used within a societal context; therefore, the societal dimension also has to be explored — leading to NBIC**S**.

Estimated Implications

If the vision is fulfilled (and nothing indicates that the vision is not feasible), we should see a drop in unemployment of disabled people. A Canadian survey found the following three accommodations are most often identified by people with disabilities not in the labor force as being necessary for them to work: (1) modified/reduced hours (33%); (2) job redesign (27%); and (3) accessible transportation (14%). The above NBICS vision should help with the elimination of these three obstacles.

If the vision is fulfilled, we also should see an increase in the level of education and knowledge of disabled people (which in itself should translate into higher employment numbers). Higher levels of knowledge and employment would lead to higher income, and that would lead to better health. Thus, NBICS would lead to better integration of disabled people into society, making them more mobile and increasing their self-esteem. The disabled, including many elderly people, will feel less isolated and will participate more in society, which will lead to many other effects, including increased well-being.

Reference

Shahriar, Lloyd S, MS, Shapiro JH, Hemmer PR. Phys Rev Lett 2001 Oct 15;87(16):167903 Long Distance, Unconditional Teleportation of Atomic States via Complete Bell State Measurements

Unison In 1998: A Canadian Approach to Disability Issues A Vision Paper Federal/Provincial/Territorial Ministers Responsible for Social Services.

D. ENHANCING GROUP AND SOCIETAL OUTCOMES

THEME SUMMARY

Panel: J.S. Albus, W.S. Bainbridge, J. Banfield, M. Dastoor, C.A. Murray, K. Carley, M. Hirshbein, T. Masciangioli, T. Miller, R. Norwood, R. Price, P. Rubin, J. Sargent, G. Strong, W.A. Wallace

The third multidisciplinary theme is concerned with NBIC innovations whose benefits would chiefly be beyond the individual level, for groups, the economy, culture, or society as a whole. It naturally builds on the human cognition and physical capabilities themes and provides a background for the national security and scientific unification panels. In particular, it is focused on a nexus issue that relates logically to most technological applications discussed in this report and that connects all four NBIC scientific and technological realms — that is, how to enhance human communication and cooperation.

The starting point for enhancing group and societal outcomes was the workshop *Societal Implications of Nanoscience and Nanotechnology*, convened by the National Science Foundation September 28-29, 2000. Members of the 2001 workshop were all given copies of the earlier workshop report (Roco and Bainbridge 2001), and they considered how to build on the earlier nanotechnology foundation to develop a broader vision giving equal weight to biotechnology, information technology, and cognitive science, with a focus on enhancing human performance.

The report of the 2000 workshop stressed that the study of the societal implications of nanotechnology must be an integral part of the National Nanotechnology Initiative, and the same is true for future NBIC efforts. The term *societal implications* refers not merely to the impact of technology on society, but also to the myriad ways in which social groups, networks, markets, and institutions may shape development of the technology. Also, as the report recognized, "...sober, technically competent research on the interactions between nanotechnology and society will help mute speculative hype and dispel some of the unfounded fears that sometimes accompany dramatic advances in scientific understanding" (Roco and Bainbridge 2001, v). Similarly, involvement of the social and behavioral sciences in the convergence of NBIC disciplines will help maximize the gains that can be achieved in human performance.

Participants first considered a wide range of likely group and societal benefits of NBIC convergence, then developed the specific vision that they judge has the greatest potential and requires the most concentrated scientific effort to achieve.

There are many potential society-wide benefits of NBIC. Working together, the NBIC sciences and technologies can increase American productivity sufficiently to maintain U.S. world leadership, solve the Social Security shortfall, and eventually eliminate poverty in the nation. NBIC can significantly help us proactively deal with the environment, create new energy sources that will reduce our reliance on foreign oil, and ensure the sustainability of our economy. Multidisciplinary research could develop a secure national integrated data system for health data that relies on nano-bio interfaces to obtain, update, and monitor personal data. Combined with new treatments and preventive measures based on NBIC convergence, such a system will extend life and improve its quality. NBIC industries of the future will employ distributed manufacturing, remote design, and production management for individualized products; cognitive control through simulated human intelligence; and a host of other techniques that will promote progress. In addition, converging technologies promise advances in simultaneous group interaction by using cognitive engineering and other new strategies.

In the vast array of very significant potential benefits of NBIC, one stands out that would catalyze all the others and that would require a special, focused effort to achieve success in the 10-20-year time frame. The panel strongly asserted that work should begin now to create *The Communicator*, a mobile system designed to enhance group communication and overcome barriers that currently prevent people from cooperating effectively. A concentrated effort involving nanotechnology, biotechnology, information technology, and cognitive science could develop in one or two decades a mature system to revolutionize people's capability to work together regardless of location or context.

The Communicator: Enhancing Group Communication, Efficiency, and Creativity

The Communicator is envisioned as a multifaceted system relying on the development of convergent technologies to enhance group communication in a wide variety of situations, including formal business or government meetings, informal social interaction, on the battlefield, and in the classroom. This system will rely on expected advances in nanotechnology fabrication and emerging information technologies, tightly coupled with knowledge obtained from the biological and cognitive domains. The convergence of these technologies will enhance individual attributes and remove barriers to group communication such as incompatible communication technologies, users' physical disabilities, language differences, geographic distance, and disparity in knowledge possessed by group members.

At the heart of The Communicator will be nano/info technologies that let individuals carry with them information about themselves and their work that can be easily shared in group situations. Thus, each individual participant will have the option to add information to the common pool of knowledge, across all domains of human experience — from practical facts about a joint task, to personal feelings about the issues faced by the group, to the goals that motivate the individual's participation.

The Communicator will also be a facilitator for group communication, an educator or trainer, and/or a translator, with the ability to tailor its personal appearance, presentation style, and activities to group and individual needs. It will be able to operate in a variety of modes, including instructor-to-group and peer-to-peer interaction, with adaptive avatars that are able to change their affective behavior to fit not only individuals and groups, but also varying situations. It will operate in multiple modalities, such as sight and sound, statistics and text, real and virtual circumstances, which can be selected and combined as needed in different ways by different participants. Improving group interactions via brain-to-brain and brain-machine-brain interactions will also be explored.

In total, a Communicator system with these attributes will be able to help overcome inequality between people, isolation of the individual from the environment, injustice and deprivation, personal and cultural biases, misunderstanding, and unnecessary conflict. In the broadest sense, it will be a powerful enhancer of communication and creativity, potentially of great economic and social benefit.

Statements and Visions

The collective vision, called The Communicator here, draws together numerous applications and sciences. In particular, it connects cognitive science and the individual-centered behavioral sciences to the broad range of group-centered social sciences. In addition, this chapter includes a vision for future transport aircraft. Thus, the statements and visions contributed by members of this working group naturally include social and well as behavioral science approaches and form a bridge back to the Roco and Bainbridge 2001 report on the societal implications of nanotechnology.

Reference

Roco, M.C. and W.S. Bainbridge, eds. 2001. *Societal Implications of Nanoscience and Nanotechnology*. Dordrecht, Netherlands: Kluwer.

STATEMENTS

COGNITION, SOCIAL INTERACTION, COMMUNICATION, AND CONVERGENT TECHNOLOGIES

Philip Rubin, National Science Foundation[1]

I am impressed with how my teenaged daughter and her friends marshal current technology for group communication. Most of their use of this technology, including AOL "Instant Messaging," email, cellphones, and transportation, is for social interaction.

The technological world twenty years from now will be a very different one. Prognostication is not my specialty and seems like a dangerous enterprise; however, I can talk about some things that we can do to help shape our future. Some of these are merely extensions of current technology and our current abilities, but the critical considerations I want to mention are well beyond our current capabilities. The unifying vision for these comments is the merging of cognition and communication.

Imagine a future without cellphones, laptops, PDAs, and other cumbersome devices. Going beyond the existing smart environments described by Reg Golledge and his colleagues (see Golledge essay in Chapter B and Loomis essay in Chapter C), we will soon be moving through a world in which we are continuously broadcasting, receiving, storing, synthesizing, and manipulating information. We will be embedded in dynamic, continually changing communicative clouds of data signals that communicate information about place, location, language, identity, persona, meaning, and intent. How will social and personal interaction be restructured in this new world? How can we use cognition to help us fly through these clouds effectively? I will leave the first question to experts like Sherry Turkle (see essay in Chapter B), who have thought long and hard about them, and will, instead, briefly mention where we need to go in the area of cognition.

The approaches that we will use for social and group communication in twenty years will rely on a variety of cognitive considerations. Here is a partial listing.

- **Intent.** *Neuro-nano* technology, such as *neural interfaces*, will enable us to provide the direct guidance of choice and selection of behaviors based on cognitive intent. This will allow for binary and graded choice directly under cognitive control.

- **Adaptation**. Communication and knowledge systems will learn and adapt based upon an understanding of human behavior. Fundamental to this is a serious consideration of the adaptive landscapes that characterize this new communicative, social world and how they mesh with our cognitive capabilities.

- **Perception, analysis, and action**. Embedded and distributed systems and sensors will be enhanced by our fundamental understanding of human perceptual and analytic behavior and skills, including the following: auditory and visual scene analysis (Biederman 1995; Bregman 1994); the visual control of action (Loomis and Beall 1998; Turvey and Remez 1979; and Warren 1988); multimodality, including vision, audition, gesture, and haptic sensing and manipulation (Cassell et al. 2000; and Turvey 1996); spatial cognition (Golledge 1999); linguistic analysis, including

[1] The views expressed in this essay do not necessarily represent the views of the National Science Foundation.

statistically-based natural language processing and analysis (Biber, Conrad, and Reppen 1998; and Manning and Schutze 1999); and language use (Clark 1996).

- **Selection**. Cognitive selection, prioritization, and organization of information are essential if the information/communication clouds of the future are not to overwhelm us. Critical abilities to filter, organize, restrict, or enhance information will rely on cognitive selection, personal preference, and automatic adaptation that will evolve based on previous behavior, patterns, choices, and preferences.

- **Semantics**. Meaning will guide the performance of the systems of the future; it will be grounded by a variety of factors, including ties to the real world and its structure and requirements, biases, and personal and social needs. Semantically based systems will make communication more flexible, effective, and natural.

- **Self-organization and complexity.** Increasingly, approaches to understanding human cognition, perception, and behavior will rely on more sophisticated analytic, statistical, and conceptual tools. Examples include nonlinear dynamical systems; self-organization, complexity and emergent behavior; complex adaptive systems; agent-based modeling; naturalistic Bayesian-networks that include subjectively-based categorization and representation; and the like (Holland 1995; Kauffman 1995, 2000; Kelso 1997; Varela et al. 1991; and Waldrop 1992. See also essay by J. Pollack in Chapter B).

What is needed to make these changes happen? First, they rely on the presumed convergence of nano-, bio-, info-, and cognitive technologies. Obviously, some of these changes are already on the way, particularly in the realm of nanotechnology, information technology, communication systems, and engineering. Progress has been significantly slower on the cognitive end, for a variety of reasons. The problems to be tackled in areas such as cognition and perception are often broad and very complex. These difficulties have been compounded by the need for noninvasive approaches for probing and exploring the human cognitive system. Mind and behavior have usually been explored from the outside. In essence, the cognitive system has been treated as a “black box” that can be probed in a variety of ways. Often such approaches have been conducted independent of the constraints imposed both by human physiology and by the environment. Other techniques that are more invasive, such as lesion studies, work with a system that is not in its normal functioning state. The difficulties in probing this system hamper our understanding of it.

Recent technological advances have raised the possibility of obtaining additional data about neural functioning during normal cognitive activities that can help to inform and constrain our theorizing. New advances in functional neuroimaging, including fMRI, PET, and MEG, coupled with the detailed study of neural circuitry and the theoretical advances in a number of areas, hold great promise (Gazzaniga et al. 1998; Lyon and Rumsey 1996; Marantz et al. 2000; and Posner and Raichle 1997). Functional imaging has the potential to be the telescope that lets us observe the universe of the mind. The goal is not to localize behavior but to have a tool that can potentially aid in the understanding of a massively complex system and in exploring brain behavior. However, these techniques will not be adequate on their own. They must be used in the context of a basic understanding of human cognition, perception, learning, development, and so forth.

Unfortunately, the fundamental understanding of how cognition works in areas such as spatial and cognition perception (auditory, haptic, and visual) has been massively underestimated. These are complex problems that will require significant basic research. For example, we need to understand our interaction with the world before we can fully understand the role the brain plays in helping us navigate this world. Before we can fully understand the role of the brain in vision, we must have a better depiction of what is available in the world for us to see. Before we fully understand the role of the brain in language, we need a clear theoretical understanding of what language is, how it is structured and organized at a variety of levels. Considerable progress that has been made in areas such as these points to the promise of theory-based research coupled with emerging technologies for visualization and simulation.

The “intelligent” systems of the future that will be fundamental to group and social communication will be far removed from the expert systems and the ungrounded formal systems of the artificial intelligence (AI) of past years. Instead, they will rely on the gains made in the fundamental understanding of the psychology, biology, and neuroscience of human behavior and performance, including cognition, perception, action, emotion, motivation, multimodality, spatial and social cognition, adaptation, linguistic analysis, and semantics. These gains will be enhanced by consideration of human behavior as a complex adaptive biological system tightly coupled to its physical and social environment.

It remains to be seen whether the national support is forthcoming that is necessary to make substantial progress in these areas of cognition that hold such promise. However, if we hope to see truly convergent technologies leading to smart devices and the enhancement of human behavior, communication, and quality of life, we must tackle the difficult problems related to cognition on the large scale more commonly seen in areas such as computer science and engineering. Now is the time to seriously begin this effort.

References

Biber, D., S. Conrad, and R. Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biederman, I. 1995. Visual object recognition. Chapter 4 in *An invitation to cognitive science, 2nd ed., Vol. 2, Visual cognition,* S.M. Kosslyn and D.N. Osherson, eds. Cambridge, MA: MIT Press.

Bregman, A.S. 1994. *Auditory scene analysis*. Cambridge, MA: MIT Press.

Cassell, J., J. Sullivan, S. Prevost, and E. Churchill. 2000. *Embodied conversational agents*. Cambridge, MA: MIT Press.

Clark, H.H. 1996. *Using language*. Cambridge: Cambridge University Press.

Gazzaniga, M.S., R.B. Ivry, and G.R. Mangun. 1998. *Cognitive neuroscience: The biology of the mind*. New York: W.W. Norton and Company.

Golledge, R.G., ed. 1999. *Wayfinding behavior: Cognitive mapping and other spatial processes*. Baltimore, MD: John Hopkins University Press.

Holland, J.H. 1995. *Hidden order: How adaptation builds complexity*. New York: Addison-Wesley.

Kauffman, S. 1995. *At home in the universe: The search for the laws of self-organization and complexity*. Oxford: Oxford University Press.

_____. 2000. *Investigations*. Oxford: Oxford University Press.

Kelso, J., and A. Scott. 1997. *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Loomis, J.M. and A. Beall. 1998. Visually controlled locomotion: Its dependence on optic flow, three-dimensional space perception, and cognition. *Ecological Psychology* 10:271-285.

Lyon, G.R. and J.M. Rumsey. 1996. *Neuroimaging: A window to the neurological foundations of learning and behavior in children*. Baltimore, MD: Paul H. Brookes Publishing Co.

Manning, C.D., and H. Schutze. 1999. *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Marantz, A., Y. Miyashita, and W. O’Neil, eds. 2000. *Image, language, brain*. Cambridge, MA: MIT Press.

Posner, M.I. and M.E. Raichle. 1997. *Images of mind*. New York: W.H. Freeman and Co.

Turvey, M.T. 1996. Dynamic touch. *American Psychologist* 51:1134-1152.

Turvey, M.T. and R.E. Remez. 1970. Visual control of locomotion in animals: An overview. In *Interrelations of the communicative senses*, L. Harmon, Ed. Washington, D.C.: National Science Foundation.

Varela, F.J., E. Thompson, and E. Rosch. 1991. *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Waldrop, M.M 1992. *Complexity: The emerging science at the edge of order and chaos*. New York: Simon and Schuster.

Warren, W.H. 1988. Action modes and laws of control for the visual guidance of action. In *Complex movement behaviour: The motor-action controversy*, O.G. Meijer and K. Roth, eds. Amsterdam: North-Holland .

ENGINEERING THE SCIENCE OF COGNITION TO ENHANCE HUMAN PERFORMANCE

William A. Wallace, Rensselaer Polytechnic Institute

The purpose of this paper is to provide a rationale for a new program whose purpose would be the integration of the science of cognition with technology to improve the performance of humans. We consider *cognition* to be “thinking” by individuals and, through consideration of emergent properties, “thinking” by groups, organizations, and societies. *Technology* is all the means employed by a social group to support its activities, in our case, to improving human performance. *Engineering* is the creation of artifacts such as technologies. Therefore, research concerned with engineering the science of cognition to improve human performance means research on the planning, design, construction, and implementation of technologies.

The purpose of such research should be to enhance performance, i.e., goal-directed behavior in a task environment, across all four levels of cognition: individual, group, organization, and society. In order to do so, we must consider the effective *integration* of cognition and technology as follows:

- integration of technology into the human central nervous system
- integration of important features of human cognition into machines
- integration of technologies (cognitive prosthetics) into the task environment to enhance human performance.

We see a synergistic combination of convergent technologies as starting with cognitive science (including cognitive neuroscience) since we need to understand the how, why, where, and when of thinking at all four levels in order to plan and design technology. Then we can employ nanoscience and nanotechnology to build the technology, and biotechnology and biomedicine to implement it. Finally, we can employ information technology to monitor and control the technology, making it work.

ENGINEERING OF MIND TO ENHANCE HUMAN PRODUCTIVITY

James S. Albus, National Institute of Standards and Technology

We have only just entered an era in history in which technology is making it possible to seriously address scientific questions regarding the nature of mind. Prior to about 125 years ago, inquiry into the nature of mind was confined to the realm of philosophy. During the first half of the 20th century, the study of mind expanded to include neuroanatomy, behavioral psychology, and psychoanalysis. The last fifty years have witnessed an explosion of knowledge in neuroscience and computational theory. The 1990s, in particular, produced an enormous expansion of understanding of the molecular and cellular processes that enable computation in the neural substrate, and more is being learned, at a faster rate, than almost anyone can comprehend:

- Research on mental disease and drug therapy has led to a wealth of knowledge about the role of various chemical transmitters in the mechanisms of neurotransmission.
- Single-cell recordings of neural responses to different kinds of stimuli have shown much about how sensory information is processed and muscles are controlled.
- The technology of brain imaging is now making it possible to visually observe where and when specific computational functions are performed in the brain.
- Researchers can literally see patterns of neural activity that reveal how computational modules work together during the complex phenomena of sensory processing, world modeling, value judgment, and behavior generation.
- It has become possible to visualize what neuronal modules in the brain are active when people are thinking about specific things, and to observe abnormalities that can be directly related to clinical symptoms (Carter 1998).

The Brain and Artificial Intelligence

In parallel developments, research in artificial intelligence and robotics has produced significant results in planning, problem-solving, rule-based reasoning, image analysis, and speech understanding. All of the fields below are active, and there exists an enormous and rapidly growing literature in each of these areas:

- Research in learning automata, neural nets, fuzzy systems, and brain modeling is providing insights into adaptation and learning and knowledge of the similarities and differences between neuronal and electronic computing processes.
- Game theory and operations research have developed methods for decision-making in the face of uncertainty.
- Genetic algorithms and evolutionary programming have developed methods for getting computers to generate successful behavior without being explicitly programmed to do so.
- Autonomous vehicle research has produced advances in realtime sensory processing, world modeling, navigation, path planning, and obstacle avoidance.
- Intelligent vehicles and weapons systems are beginning to perform complex military tasks with precision and reliability.

- Research in industrial automation and process control has produced hierarchical control systems, distributed databases, and models for representing processes and products.

- Computer-integrated manufacturing research has achieved major advances in the representation of knowledge about object geometry, process planning, network communications, and intelligent control for a wide variety of manufacturing operations.

- Modern control theory has developed precise understanding of stability, adaptability, and controllability under various conditions of uncertainty and noise.

- Research in sonar, radar, and optical signal processing has developed methods for fusing sensory input from multiple sources, and assessing the believability of noisy data.

In the field of *software engineering*, progress is also rapid, after many years of disappointing results. Much has been learned about how to write code for software agents and build complex systems that process signals, understand images, model the world, reason and plan, and control complex behavior. Despite many false starts and overly optimistic predictions, artificial intelligence, intelligent control, intelligent manufacturing systems, and smart weapons systems have begun to deliver solid accomplishments:

- We are learning how to build systems that learn from experience, as well as from teachers and programmers.

- We understand how to use computers to measure attributes of objects and events in space and time.

- We know how to extract information, recognize patterns, detect events, represent knowledge, and classify and evaluate objects, events, and situations.

- We know how to build internal representations of objects, events, and situations, and how to produce computer-generated maps, images, movies, and virtual reality environments.

- We have algorithms that can evaluate cost and benefit, make plans, and control machines.

- We have engineering methods for extracting signals from noise.

- We have solid mathematical procedures for making decisions amid uncertainty.

- We are developing new manufacturing techniques to make sensors tiny, reliable, and cheap.

- Special-purpose integrated circuits can now be designed to implement neural networks or perform parallel operations such as are required for low-level image processing.

- We know how to build human-machine interfaces that enable close coupling between humans and machines.

- We are developing vehicles that can drive without human operators on roads and off.

- We are discovering how to build controllers that generate autonomous tactical behaviors under battlefield conditions.

As the fields of brain research and intelligent systems engineering converge, the probability grows that we may be able to construct what Edelman (1999) calls a “conscious artifact.” Such a development would provide answers to many long-standing scientific questions regarding the relationship between the mind and the body. At the very least, building artificial models of the mind would provide new insights into mental illness, depression, pain, and the physical bases of perception, cognition, and behavior. It would open up new lines of research into questions that hitherto have not been amenable to scientific investigation:

- We may be able to understand and describe intentions, beliefs, desires, feelings, and motives in terms of computational processes with the same degree of precision that we now can apply to the exchange of energy and mass in radioactive decay or to the sequencing of amino acid pairs in DNA.
- We may discover whether humans are unique among the animals in their ability to have feelings, and start to answer the questions,
  - To what extent do humans alone have the ability to experience pain, pleasure, love, hate, jealousy, pride, and greed?
  - Is it possible for artificial minds to appreciate beauty and harmony or comprehend abstract concepts such as truth, justice, meaning, and fairness?
  - Can silicon-based intelligence exhibit kindness or show empathy?
  - Can machines pay attention, be surprised, or have a sense of humor?
  - Can machines feel reverence, worship God, be agnostic?

Engineering Intelligent Systems

The book *Engineering of Mind: An Introduction to the Science of Intelligent Systems* (Albus and Meystel 2001) outlines the main streams of research that we believe will eventually converge in a scientific theory that can support and bring about the engineering of mind. We believe that our research approach can enable the design of intelligent systems that pursue goals, imagine the future, make plans, and react to what they see, feel, hear, smell, and taste. We argue that highly intelligent behavior can be achieved by decomposing goals and plans through many hierarchical levels, with knowledge represented in a world model at the appropriate range and resolution at each level. We describe how a high degree of intelligence can be achieved using a rich dynamic world model that includes both *a priori* knowledge and information provided by sensors and a sensory processing system. We suggest how intelligent decision-making can be facilitated by a value judgment system that evaluates what is good and bad, important and trivial, and one that estimates cost, benefit, and risk of potential future actions. This will enable the development of systems that behave as if they are sentient, knowing, caring, creative individuals motivated by hope, fear, pain, pleasure, love, hate, curiosity, and a sense of priorities.

We believe that this line of research on highly intelligent systems will yield important insights into elements of mind such as attention, gestalt grouping, filtering, classification, imagination, thinking, communication, intention, motivation, and subjective experience. As the systems we build grow increasingly intelligent, we will begin to see the outlines of what can only be called mind. We hypothesize that mind is a phenomenon that will emerge when intelligent systems achieve a certain level of sophistication in sensing, perception, cognition, reasoning, planning, and control of behavior.

There are good reasons to believe that the computing power to achieve human levels of intelligence will be achieved within a few decades. Since computers were invented about a half-century ago, the rate of progress in computer technology has been astounding. Since the early 1950s, computing power has doubled about every three years. This is a compound growth rate of a factor of ten per decade, a factor of 100 every two decades. This growth rate shows no sign of slowing, and in fact, is accelerating: during the 1990s, computing power doubled every 18 months — a factor of ten every five years. Today, a typical personal computer costing less than $1000 has more computing power than a top-of-the-line supercomputer of only two decades ago. One giga-op (one billion operations per second) single-board computers are now on the market. There appears to be no theoretical limit that will slow the rate of growth in computing power for at least the next few decades. This means that within ten years, a relatively inexpensive network of ten single-board computers could have computational power approaching one tera-ops (one trillion, or 10^{12} operations per second). Within twenty years, ten single-board computers will be capable of 10^{14} operations per second. This is equivalent to the estimated computational power of the human brain (Moravec 1999). Thus, it seems quite likely that within two decades, the computing power will exist to build machines that are functionally equivalent to the human brain.

Of course, more than raw computing power is necessary to build machines that achieve human levels of performance. But the knowledge of how to utilize this computing power to generate highly intelligent behavior is developing faster than most people appreciate. Progress is rapid in many different fields. Recent results from a number of disciplines have established the foundations for a theoretical framework that might best be called a "computational theory of mind." In our book, Meystel and I have organized these results into a reference model architecture that we believe can be used to organize massive amounts of computational power into intelligent systems with human-level capabilities. This reference model architecture consists of a hierarchy of massively parallel computational modules and data structures interconnected by information pathways that enable analysis of the past, estimation of the present, and prediction of the future.

This architecture specifies a rich dynamic internal model of the world that can represent entities, events, relationships, images, and maps in support of higher levels of intelligent behavior. This model enables goals, motives, and priorities to be decomposed into behavioral trajectories that achieve or maintain goal states. Our reference architecture accommodates concepts from artificial intelligence, control theory, image understanding, signal processing, and decision theory. We demonstrate how algorithms, procedures, and data embedded within this architecture can enable the analysis of situations, the formulation of plans, the choice of behaviors, and the computation of current and expected rewards, punishments, costs, benefits, risks, priorities, and motives.

Our reference model architecture suggests an engineering methodology for the design and construction of intelligent machine systems. This architecture consists of layers of interconnected computational nodes, each containing elements of sensory processing, world modeling, value judgment, and behavior generation. At lower levels, these elements generate goal-seeking reactive behavior; at higher levels, they enable perception, cognition, reasoning, imagination, and planning. Within each level, the product of range and resolution in time and space is limited: at low levels, range is short and resolution is high, whereas at high levels, range is long and resolution is low. This enables high precision and quick response to be achieved at low levels over short intervals of time and space, while long-range plans and abstract concepts can be formulated at high levels over broad regions of time and space.

Our reference model architecture is expressed in terms of the Realtime Control System (RCS) that has been developed at the National Institute of Standards and Technology and elsewhere over the last 25 years. RCS provides a design methodology, software development tools, and a library of software that is free and available via the Internet. Application experience with RCS provides examples of how this reference model can be applied to problems of practical importance. As a result of this experience, we believe that the engineering of mind is a feasible scientific goal that could be achieved within the next quarter century.

Implications for the Future

Clearly, the ability to build highly intelligent machine systems will have profound implications — in four important areas in particular: science, economic prosperity, military power, and human wellbeing, as detailed below.

D. Science

All of science revolves around three fundamental questions:

1. What is the nature of matter and energy?

2. What is the nature of life?

3. What is the nature of mind?

Over the past 300 years, research in the physical sciences has produced a wealth of knowledge about the nature of matter and energy, both on our own planet and in the distant galaxies. We have developed mathematical models that enable us to understand at a very deep level what matter is, what holds it together, and what gives it its properties. Our models of physics and chemistry can predict with incredible precision how matter and energy will interact under an enormous range of conditions. We have a deep understanding of what makes the physical universe behave as it does. Our knowledge includes precise mathematical models that stretch over time and space from the scale of quarks to the scale of galaxies.

Over the past half-century, the biological sciences have produced a revolution in knowledge about the nature of life. We have developed a wonderfully powerful model of the molecular mechanisms of life. The first draft of the human genome has been published. We may soon understand how to cure cancer and prevent AIDS. We are witnessing an explosion in the development of new drugs and new sources of food. Within the next century, biological sciences may eliminate hunger, eradicate most diseases, and discover how to slow or even reverse the aging process.

Yet, of the three fundamental questions of science, the most profound may be, “What is mind?” Certainly this is the question that is most relevant to understanding the fundamental nature of human beings. We share most of our body chemistry with all living mammals. Our DNA differs from that of chimpanzees by only a tiny percentage of the words in the genetic code. Even the human brain is similar in many respects to the brains of apes. Who we are, what makes us unique, and what distinguishes us from the rest of creation lies not in our physical elements, or even in our biological make up, but in our minds.

It is only the mind that sharply distinguishes the human race from all the other species. It is the mind that enables humans to understand and use language, to manufacture and use tools, to tell stories, to compute with numbers, and reason with rules of logic. It is the mind that enables us to compose music and poetry, to worship, to develop technology, and organize political and religious institutions. It is the mind that enabled humans to discover how to make fire, to build a wheel, to navigate a ship, to smelt copper, refine steel, split the atom, and travel to the moon.

The mind is a process that emerges from neuronal activity within the brain. The human brain is arguably the most complex structure in the known universe. Compared to the brain, the atom is an uncomplicated bundle of mass and energy that is easily studied and well understood. Compared to the brain, the genetic code embedded in the double helix of DNA is relatively straightforward. Compared to the brain, the molecular mechanisms that replicate and retrieve information stored in the genes are quite primitive. One of the greatest mysteries in science is how the computational mechanisms in the brain generate and coordinate images, feelings, memories, urges, desires, conceits, loves, hatreds, beliefs, pleasures, disappointment, and pain that make up human experience. The really great scientific question is "What causes us to think, imagine, hope, fear, dream, and act like we do?" Understanding the nature of mind may be the most interesting and challenging problem in all of science.

Economic Prosperity

Intelligent machines can and do create wealth. And as they become more intelligent, they will create more wealth. Intelligent machines will have a profound impact on the production of goods and services. Until the invention of the computer, economic wealth (i.e., goods and services) could not be generated without a significant amount of human labor (Mankiw 1992). This places a fundamental limit on average per capita income. Average income cannot exceed average worker productive output. However, the introduction of the computer into the production process is enabling the creation of wealth with little or no human labor. This removes the limit to average per capita income. It will almost certainly produce a new industrial revolution (Toffler 1980).

The first industrial revolution was triggered by the invention of the steam engine and the discovery of electricity. It was based on the substitution of mechanical energy for muscle power in the production of goods and services. The first industrial revolution produced an explosion in the ability to produce material wealth. This led to the emergence of new economic and political institutions. A prosperous middle class based on industrial production and commerce replaced aristocracies based on slavery. In all the thousands of centuries prior to the first industrial revolution, the vast majority of humans existed near the threshold of survival, and every major civilization was based on slavery or serfdom. Yet, less than three hundred years after the beginning of the first industrial revolution, slavery has almost disappeared, and a large percentage of the World's population lives in a manner that far surpasses the wildest utopian fantasies of former generations.

There is good reason to believe that the next industrial revolution will change human history at least as profoundly as the first. The application of computers to the control of industrial processes is bringing into being a new generation of machines that can create wealth largely or completely unassisted by human beings. The next industrial revolution, sometimes referred to as the robot revolution, has been triggered by the invention of the computer. It is based on the substitution of electronic computation for the human brain in the control of machines and industrial processes. As intelligent machine systems become more and more skilled and numerous in the production process, productivity will rise and the cost of labor, capital, and material will spiral downward. This will have a profound impact on the structure of civilization. It will undoubtedly give rise to new social class structures and new political and economic institutions (Albus 1976).

The Role of Productivity

The fundamental importance of productivity on economic prosperity can be seen from the following equation:

Output = Productivity x Input

where

Input = labor + capital + raw materials and

Productivity = the efficiency by which the input of labor, capital, and raw material is transformed into output product

Productivity is a function of knowledge and skill, i.e., technology. Growth in productivity depends on improved technology. The rapid growth in computer technology has produced an unexpectedly rapid increase in productivity that has confounded predictions of slow economic growth made by establishment economists only a decade ago (Symposia 1988; Bluestone and Harrison 2000). In the future, the introduction of truly intelligent machines could cause productivity to grow even faster. Given only conservative estimates of growth in computer power, unprecedented rates of productivity growth could become the norm as intelligent machines become pervasive in the productive process.

Intelligent systems have the potential to produce significant productivity improvements in many sectors of the economy, both in the short term and in the long term. Already, computer-controlled machines routinely perform economically valuable tasks in manufacturing, construction, transportation, business, communications, entertainment, education, waste management, hospital and nursing support, physical security, agriculture and food processing, mining and drilling, and undersea and planetary exploration.

As intelligent systems become widespread and inexpensive, productivity will grow and the rate of wealth production will increase. Intelligent machines in manufacturing and construction will increase the stock of wealth and reduce the cost of material goods and services. Intelligent systems in health care will improve services and reduce costs for the sick and elderly. Intelligent systems could make quality education available to all. Intelligent systems will make it possible to clean up and recycle waste, reduce pollution, and create environmentally friendly methods of production and consumption.

The potential impact of intelligent machines is magnified by that fact that technology has reached the point where intelligent machines have begun to exhibit a capacity for self-reproduction. John von Neumann (1966) was among the first to recognize that machines can possess the ability to reproduce. Using mathematics of finite state machines and Turing machines, von Neumann developed a theoretical proof that machines can reproduce. Over the past two decades, the theoretical possibility of machine reproduction has been empirically demonstrated (at least in part) in the practical world of manufacturing:

- Computers are routinely involved in the processes of manufacturing computers
- Computers are indispensable to the process of designing, testing, manufacturing, programming, and servicing computers
- On a more global scale, intelligent factories build components for intelligent factories

At a high level of abstraction, many of the fundamental processes of biological and machine reproduction are similar. Some might object to a comparison between biological and machine reproduction on the grounds that the processes of manufacturing and engineering are fundamentally different from the processes of biological reproduction and evolution. Certainly there are many essential differences between biological and machine reproduction. But the comparison is not entirely far-fetched. And the results can be quite similar. Both biological and machine reproduction can produce populations that grow exponentially. In fact, machine reproduction can be much faster than biological. Intelligent machines can flow from a production line at a rate of many per hour.

Perhaps more important, machines can evolve from one generation to the next much faster and more efficiently than biological organisms. Biological organisms evolve by a Darwinian process, through random mutation and natural selection. Intelligent machines evolve by a Lamarckian process, through conscious design improvements under selective pressures of the marketplace. In the machine evolutionary process, one generation of computers often is used to design and manufacture the next generation of more powerful and less costly computers. Significant improvements can occur in a very short time between one generation of machines and the next. As a result, intelligent machines are evolving extremely quickly relative to biological species. Improved models of computer systems appear every few months to vie with each other in the marketplace. Those that survive and are profitable are improved and enhanced. Those that are economic failures are abandoned. Entire species of computers evolve and are superceded within a single decade. In other words, machine reproduction, like biological reproduction, is subject to evolutionary pressures that tend to reward success and punish failure.

The ability of intelligent systems to reproduce and evolve will have a profound effect on the capacity for wealth production. As intelligent machines reproduce, their numbers will multiply, leading to an exponential increase in the intelligent machine population. Since intelligent machines can increase productivity and produce wealth, this implies that with each new generation of machine, goods and services will become dramatically less expensive and more plentiful, while per capita wealth will increase exponentially.

The Prospects for Technology Growth

It is sometimes argued that technology, and therefore productivity, cannot grow forever because of the law of diminishing returns. It is argued that there must be a limit to everything, and therefore, productivity cannot grow indefinitely. Whether this is true in an abstract sense is an interesting philosophical question. Whether it is true in any practical sense is clear: it is not. From the beginning of human civilization until now, it remains a fact that the more that is known, the easier it is to discover new knowledge. And there is nothing to suggest that knowledge will be subject to the law of diminishing returns in the foreseeable future. Most of the scientists who have ever lived are alive and working today. Scientists and engineers today are better educated and have better tools with which to work than ever before. In the neurological and cognitive sciences, the pace of discovery is astonishing. The same is true in computer science, electronics, manufacturing, and many other fields. Today, there is an explosion of new knowledge in almost every field of science and technology.

There is certainly no evidence that we are nearing a unique point in history where progress will be limited by an upper bound on what there is to know. There is no reason to believe that such a limit even exists, much less that we are approaching it. On the contrary, there is good evidence that the advent of intelligent machines has placed us on the cusp of a growth curve where productivity can grow exponentially for many decades, if not indefinitely. Productivity growth is directly related to growth in knowledge. Growth in knowledge is dependent on the amount and effectiveness of investment in research, development, and education. This suggests that, given adequate investment in technology, productivity growth could return to 2.5 percent per year, which is the average for the twentieth century. With higher rates of investment, productivity growth could conceivably rise to 4 percent, which is the average for the 1960-68 time frame. Conceivably, with sufficient investment, productivity growth could exceed 10 percent, which occurred during the period between 1939 and 1945 (Samuelson and Nordhaus 1989).

If such productivity growth were to occur, society could afford to improve education, clean up the environment, and adopt less wasteful forms of production and consumption. Many social problems that result from slow economic growth, such as poverty, disease, and pollution, would virtually disappear. At the same time, taxes could be reduced, Social Security benefits increased, healthcare and a minimum income could be provided for all. The productive capacity of intelligent machines could generate sufficient per capita wealth to support an aging population without raising payroll taxes on a shrinking human labor force. Over the next three decades, intelligent machines might provide the ultimate solution to the Social Security and Medicare crisis. Benefits and services for an aging population could be continuously expanded, even in countries with stable or declining populations.

Military Power

Intelligent systems technologies have the potential to revolutionize the art of war. The eventual impact on military science may be as great as the invention of gunpowder, the airplane, or nuclear weapons. Intelligent weapons systems are already beginning to emerge. Cruise missiles, smart bombs, and unmanned reconnaissance aircraft have been deployed and used in combat with positive effect. Unmanned ground vehicles and computer-augmented command and control systems are currently being developed and will soon be deployed. Unmanned undersea vehicles are patrolling the oceans collecting data and gathering intelligence. These are but the vanguard of a whole new generation of military systems that will become possible as soon as intelligent systems engineering becomes a mature discipline (Gourley 2000).

In future wars, unmanned air vehicles, ground vehicles, ships, and undersea vehicles will be able to outperform manned systems. Many military systems are limited in performance because of the inability of the human body to tolerate high levels of temperature, acceleration, vibration, or pressure, or because humans need to consume air, water, and food. A great deal of the weight and power of current military vehicles is spent on armor and life support systems that would be unnecessary if there were no human operators on board. A great deal of military tactics and strategy are based on the need to minimize casualties and rescue people from danger. This would become unnecessary if warriors could remain out of harm's way.

Intelligent military systems will significantly reduce the cost of training and readiness. Compared to humans, unmanned vehicles and weapons systems will require little training or maintenance to maintain readiness. Unmanned systems can be stored in forward bases or at sea for long periods of time at low cost. They can be mobilized quickly in an emergency, and they will operate without fear under fire, the first time and every time.

Intelligent systems also enable fast and effective gathering, processing, and displaying of battlefield information. They can enable human commanders to be quicker and more thorough in planning operations and in replanning as unexpected events occur during the course of battle. In short, intelligent systems promise to multiply the capabilities of the armed forces, while reducing casualties and hostages and lowering the cost of training and readiness (Maggart and Markunas 2000).

Human Wellbeing

It seems clear that intelligent systems technology will have a profound impact on economic growth. In the long run, the development of intelligent machines could lead to a golden age of prosperity, not only in the industrialized nations, but throughout the world. Despite the explosion of material wealth produced by the first industrial revolution, poverty persists and remains a major problem throughout the world today. Poverty causes hunger and disease. It breeds ignorance, alienation, crime, and pollution. Poverty brings misery, pain, and suffering. It leads to substance abuse. Particularly in the third world, poverty may be the biggest single problem that exists, because it causes so many other problems. And yet there is a well-known cure for poverty. It is wealth.

Wealth is difficult to generate. Producing wealth requires labor, capital, and raw materials — multiplied by productivity. The amount of wealth that can be produced for a given amount of labor, capital, and raw materials depends on productivity. The level of productivity that exists today is determined by the current level of knowledge embedded in workers' skills, management techniques, tools, equipment, and software used in the manufacturing process. In the future, the level of productivity will depend more and more on the level of knowledge embedded in intelligent machines. As the cost of computing power drops and the skills of intelligent machines grow, the capability for wealth production will grow exponentially. The central question then becomes, how will this wealth be distributed?

In the future, new economic theories based on abundance may emerge to replace current theories based on scarcity. New economic institutions and policies may arise to exploit the wealth-producing potential of large numbers of intelligent machines. As more wealth is produced without direct human labor, the distribution of income may shift from wages and salaries to dividends, interest, and rent. As more is invested in ownership of the means of production, more people may derive a substantial income from ownership of capital stock. Eventually, some form of people's capitalism may replace the current amalgam of capitalism and socialism that is prevalent in the industrialized world today (Albus 1976; Kelso and Hetter 1967).

Summary and Conclusions

We are at a point in history where science has good answers to questions such as, "What is the universe made of?" and "What are the fundamental mechanisms of life?" There exists a wealth of knowledge about how our bodies work. There are solid theories for how life began and how species evolved. However, we are just beginning to acquire a deep understanding of how the brain works and what the mind is.

We know a great deal about how the brain is wired and how neurons compute various functions. We have a good basic understanding of mathematics and computational theory. We understand how to build sensors, process sensory information, extract information from images, and detect entities and events. We understand the basic principles of attention, clustering, classification, and statistical analysis. We understand how to make decisions in the face of uncertainty. We know how to use knowledge about the world to predict the future, to reason, imagine, and plan actions to achieve goals. We have algorithms that can decide what is desirable, and plan how to get it. We have procedures to estimate costs, risks, and benefits of potential actions. We can write computer programs to deal with uncertainty and compensate for unexpected events. We can build machines that can parse sentences and extract meaning from messages, at least within the constrained universe of formal languages.

As computing power increases and knowledge grows of how the brain converts computational power into intelligent behavior, the ability of machines to produce greater wealth (i.e., goods and services that people want and need) will enable many possible futures that could never before have been contemplated. Even under very conservative assumptions, the possibilities that can be generated from simple extrapolations of current trends are very exciting. We are at a point in history where some of the deepest mysteries are being revealed. We are discovering how the brain processes information, how it represents knowledge, how it makes decisions and controls actions. We are beginning to understand what the mind is. We will soon have at our disposal the computational power to emulate many of the functional operations in the brain that give rise to the phenomena of intelligence and consciousness. We are learning how to organize what we know into an architecture and methodology for designing and building truly intelligent machines. And we are developing the capacity to experimentally test our theories. As a result, we are at the dawning of an age where the engineering of mind is feasible.

In our book (Albus and Meystel 2001), we have suggested one approach to the engineering of mind that we believe is promising, containing the following elements:

- *a perception system* that can fuse *a priori* knowledge with current experience and can understand what is happening, both in the outside world and inside the system itself

- *a world-modeling system* that can compute what to expect and predict what is likely to result from contemplated actions
- *a behavior-generating system* that can choose what it intends to do from a wide variety of options and can focus available resources on achieving its goals
- *a value judgment system* that can distinguish good from bad and decide what is desirable

We have outlined a reference model architecture for organizing the above functions into a truly intelligent system, hypothesizing that in the near future it will become possible to engineer intelligent machines with intentions and motives that use reason and logic to devise plans to accomplish their objectives.

Engineering of mind is an enterprise that will prove at least as technically challenging as the Apollo program or the Human Genome project. And we are convinced that the potential benefits for humankind will be at least as great, perhaps much greater. Understanding of the mind and brain will bring major scientific advances in psychology, neuroscience, and education. A computational theory of mind may enable us to develop new tools to cure or control the effects of mental illness. It will certainly provide us with a much deeper appreciation of who we are and what our place is in the universe.

Understanding of the mind and brain will enable the creation of a new species of intelligent machine systems that can generate economic wealth on a scale hitherto unimaginable. Within a half-century, intelligent machines might create the wealth needed to provide food, clothing, shelter, education, medical care, a clean environment, and physical and financial security for the entire world population. Intelligent machines may eventually generate the production capacity to support universal prosperity and financial security for all human beings. Thus, the engineering of mind is much more than the pursuit of scientific curiosity. It is more even than a monumental technological challenge. It is an opportunity to eradicate poverty and usher in a golden age for all human kind.

References

Albus, J.S. and A.M. Meystel. 2001. *Engineering of mind: An introduction to the science of intelligent systems.* New York: John Wiley and Sons.

Albus, J.S. 1976. *Peoples' capitalism: The economics of the robot revolution*. Kensington, MD: New World Books. See also Peoples' Capitalism web page at http://www.peoplescapitalism.org.

Bluestone, B., and B. Harrison. 2000. *Growing prosperity: The battle for growth with equity in the twenty-first century.* New York: Houghton Mifflin Co.

Carter, R. 1998. *Mapping the mind*. University of California Press.

Edelman, G. 1999. Proceedings of International Conference on Frontiers of the Mind in the 21st Century, Library of Congress, Washington D.C., June 15

Gourley, S.R. 2000. Future combat systems: A revolutionary approach to combat victory. *Army* 50(7):23-26 (July).

Kelso, L., and P. Hetter. 1967. *Two factor theory: The economics of reality*. New York: Random House.

Maggart, L.E., and R.J. Markunas. 2000. Battlefield dominance through smart technology. *Army* 50(7).

Mankiw, G.N. 1992. *Macroeconomics*. New York: Worth Publishers.

Moravec, H. 1999. *Robot: Mere machine to transcendent mind*. Oxford: Oxford University Press.

Samuelson, P., and W. Nordhaus. 1989. *Economics*, 13th ed. New York: McGraw-Hill.

Symposia. 1988. The slowdown in productivity growth. *Journal of Economic Perspectives* 2 (Fall).

Toffler, A. 1980. *The third wave*. New York: William Morrow and Co.

von Neumann, J. 1966. Theory of self-reproducing automata (edited and completed by A. Burks). Urbana: University of Illinois Press.

MAKING SENSE OF THE WORLD: CONVERGENT TECHNOLOGIES FOR ENVIRONMENTAL SCIENCE

Jill Banfield, University of California, Berkeley

Through the combination of geoscience, biology, and nano- and information technologies, we can develop a fundamental understanding of the factors that define and regulate Earth's environments from the molecular to global scale. It is essential that we capture the complex, interconnected nature of the processes that maintain the habitability of the planet in order to appropriately utilize Earth's resources and predict, monitor, and manage global change. This goal requires long-term investments in nanogeoscience, nanotechnology, and biogeochemical systems modeling.

Introduction

Looking to the future, what are the greatest challenges our society (and the world) faces? Ensuring an adequate food supply, clean air, and clean water, are problems intimately linked to the environment. Given the rate of accumulation of environmental damage, it seems appropriate to ask, can science and technology solve the problems associated with pollution and global change before it is too late? Where should we invest our scientific and technological efforts, and what might these investments yield?

One of the mysteries concerning environmental processes is the role of extremely small particles that, to date, have defied detection and/or characterization. We now realize that materials with dimensions on the nanometer scale (intermediate between clusters and macroscopic crystals) are abundant and persistent in natural systems. Nanoparticles are products of, and substrates for, nucleation and growth in clouds. They are also the initial solids formed in water, soils, and sediments. They are generated in chemical weathering and biologically mediated redox reactions, during combustion of fuel, and in manufacturing. For example, nanoparticles are by-products of microbial energy generation reactions that utilize inorganic ions (e.g., Mn, Fe, S, U) as electron donors or acceptors. They are highly reactive due to their large surface areas, novel surface structures, and size-dependent ion adsorption characteristics and electronic structures (including redox potentials). It is likely that they exert a disproportionately large, but as yet incompletely defined, influence on environmental geochemistry because they provide a means for transport of insoluble ions and present abundant novel, reactive surfaces upon which reactions, including catalytic reactions, occur.

It is widely accepted that the most rapid growth in knowledge in recent years has occurred in the field of biology. In the environmental context, the biology of single-celled organisms represents a critically important focus, for several reasons. First, microbes are extraordinarily abundant. They underpin many of the biogeochemical cycles in the environment and thus directly impact the bioavailability of contaminants and nutrients in ecosystems. They are responsible for the formation of reactive mineral particles and contribute to mineral dissolution. With analysis of these connections comes the ability to use microbes to solve environmental problems. Second, microorganisms are relatively simple, hence detailed analysis of how they work represents a tractable problem. Third, microbes have invented ways to carry out chemical transformations via enzymatic pathways at low temperatures. These pathways have enormous industrial potential because they provide energetically inexpensive routes to extract, concentrate, and assemble materials needed by society. Identification of the relevant microbial enzymatic or biosynthetic pathways requires analysis of the full diversity of microbial life, with emphasis on organisms in extreme natural geologic settings where metabolisms are tested at their limits.

Where does our understanding of microbes and nanoparticles in the environment stand today? Despite the fact that microbes dominate every habitable environment on Earth, we know relatively little about how most microbial cells function. Similarly, we have only just begun to connect the novel properties and reactivity of nanoparticles documented in the laboratory to phenomena in the environment. Although our understanding of these topics is in its infancy, science is changing quickly. The center of this revolution is the combination of molecular biology, nanoscience, and geoscience.

The Role of Converging Technologies

In order to comprehensively understand how environmental systems operate at all scales, convergence of biological, technological, and geoscientific approaches is essential. Three important tasks are described below.

Identification and Analysis of Reactive Components in Complex Natural Systems

Nanoparticles and microorganisms are among the most abundant, most reactive components in natural systems. Natural nanoparticles (often < 5 nm in diameter) exhibit the same novel size-dependent properties that make their synthetic equivalents technologically useful. The functions of some microbial cell components (e.g., cell membranes, ribosomes) probably also depend on size-related reactivity. A challenge for the immediate future is determination of the origin, diversity, and roles of nanoparticles in the environment. Similarly, it is critical that we move from detecting the full diversity of microorganisms in most natural systems to understanding their ranges of metabolic capabilities and the ways in which they shape their environments. These tasks require integrated characterization studies that provide molecular-level (inorganic and biochemical) resolution.

Massive numbers of genetic measurements are needed in order to identify and determine the activity of thousands of organisms in air, water, soils, and sediments. Enormous numbers of chemical measurements are also required in order to characterize the physical environment and to evaluate how biological and geochemical processes are interconnected. This task demands laboratory and field data that is spatially resolved at the submicron-scale at which heterogeneities are important, especially in interfacial regions where reactions are fastest. The use of robots in oceanographic monitoring studies is now standard, but this is only the beginning. Microscopic devices are needed to make *in situ*, fine-scale measurements of all parameters and to conduct *in situ* experiments (e.g., to assay microbial population makeup in algal blooms in the ocean or to determine which specific organism is responsible for biodegradation of an organic pollutant in a contaminated aquifer). These devices are also required for instrumentation of field sites to permit monitoring over hundreds of meters to kilometer-scale distances. Development of appropriate microsensors for these applications is essential.

Environmental science stands to benefit greatly from nanotechnology, especially if new sensors are developed with environmental monitoring needs in mind. In the most optimistic extreme, the sensors may be sufficiently small to penetrate the deep subsurface via submicron-scale pores and be able to relay their findings to data collection sites. It is likely that these extremely small, durable devices also will be useful for extraterrestrial exploration (e.g., Mars exploration).

Monitoring Processes in the Deep Subsurface

Many of the inorganic and organic contaminants and nutrients of interest in the environment may be sequestered at considerable depths in aquifers or geological repositories. Methods are needed to image the structure of the subsurface to locate and identify these compounds, determine the nature of their surroundings, and monitor changes occurring during natural or enhanced *in situ* remediation. Examples of problems for study include detection of nanoparticulate metal sulfide or uranium oxide minerals produced by biological reduction, possibly via geophysical methods; analysis of the role of transport of nanoparticulate contaminants away from underground nuclear waste repositories; and monitoring of the detailed pathways for groundwater flow and colloid transport.

Development of Models to Assist in Analysis of Complex, Interdependent Phenomena

After we have identified and determined the distributions of the reactive inorganic and organic nanoscale materials in natural systems, it is essential that we understand how interactions between these components shape the environment. For example, we anticipate development and validation of comprehensive new models that integrate predictions of particle-particle organic aggregation and crystal growth with models that describe how aggregates are transported through porous materials in the subsurface. These developments are essential for prediction of the transport and fate of contaminants during and after environmental remediation.

Environmental processes operate across very large scales on continents and in the oceans. Thus, remote collection of high-resolution data sets (e.g., by satellite-based remote sensing) can also be anticipated. The large quantities of data from direct and indirect monitoring programs will benefit from new methodologies for information management. Mathematical models are essential to guide cognition and to communicate the principles that emerge from the analyses. An example of an ecosystem model is shown in Figure D.1. Input from the cognitive sciences will be invaluable to guide development of supermodels of complex processes.

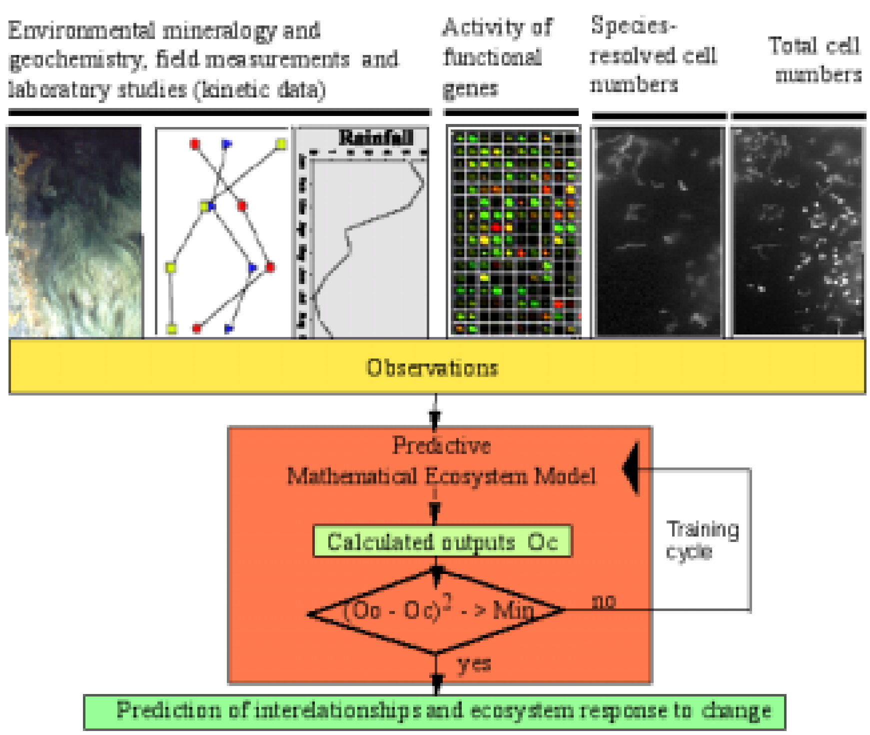


Figure D.1. Example of an ecosystem model that incorporates information about the physical and chemical environment with information about population size and structure and gene expression to analyze community interactions and predict response of the system to perturbations.

The Transforming Strategy

The first task toward an integrated understanding of the Earth's ecosystems is to identify and study the most important components. Focus on microorganisms is warranted based on their sheer abundance and metabolic versatility. The first two disciplinary partners, molecular biology and nanoscience, have already taken center stage with the integration of molecular biology and genome-enabled technologies (e.g., whole genome expression microarrays). In the next few years, these tools will allow us to decipher the full diversity of ways in which individual organisms grow, develop, reproduce, and evolve. These breakthroughs are critical to medicine, agriculture, and biologically assisted manufacturing and waste management.

Inorganic components also play key roles in natural systems. As noted above, exceedingly small particles intermediate in size between molecular clusters and macroscopic materials (nanoparticles) are abundant components of natural environments. Study of nanoparticle formation, properties, and stability is at the intersection of nanoscience, biology, chemistry, and geoscience. The unique characteristics of materials structured on the nanoscale have long been appreciated in the fields of materials science and engineering. It is now essential that we determine whether the nanoparticles in soils, sediments, water, the atmosphere, and in space also have unusual and environmentally important surface properties and reactivity. Do nanoparticles partition in natural systems in size-dependent ways? Are they transported readily in groundwater, and is this the mechanism by which insoluble contaminants and nutrients are dispersed? There are also potentially intriguing questions relating to interactions between inorganic nanoparticles and organic molecules. For example, do nanoparticles in dust react in unusual ways with organic molecules (perhaps in sunlight)? Is the assembly of nanoparticles by organic polymers central to biomineralization processes, such as generation of bone? Can these interactions be harnessed for biomimetic technologies? Did reactions at nanoparticle surfaces play a role in prebiotic synthesis or the origin of life? Were nanoparticles themselves captured by organic molecules to form early enzymes? The answers to these questions are important to our understanding of inorganic and biological systems. However, far larger challenges remain.

The second task will be to investigate entire communities of microorganisms at the genetic level to provide new insights into community structure and organization, including cell-cell signaling and the partitioning of function. This challenge requires complete genetic analysis of all community members without cultivation. This task will require extension of current biological, computational, and information technologies to permit simultaneous reconstruction of genome content from multiorganism assemblages at the level of strains without isolation of each community member. Resulting data will also allow comparison of the microbial community lifestyle — characterized by the ability to directly control the geochemical cycles of virtually every element — to its alternative, multicellular life. These analyses will also unveil the pathways by which all biologically and geochemically important transformations are accomplished. This work must be initiated in the laboratory, but ultimately, must be expanded to explicitly include all environmental parameters and stimuli. Consequently, the task of understanding organisms in their environments stands before us as the third and longest-term task.

An additional component, geoscience, must be included in order to meet the challenge of molecularly resolved ecology. Environmental applications have lagged behind investigations of organisms in the laboratory because natural systems are extremely complicated. Critical environmental data include time-resolved measurements of the structure and organization of natural systems, organism population statistics, measurements of the levels of expression of all genes within communities of interacting species, and quantification of how these expression patterns are controlled by and control geochemical processes. This approach, which must ultimately include macroorganisms, will be essential for medical and agricultural, as well as environmental, reasons.

Education and Outreach

Analysis of complex systems, through integration of nanotechnology, nanoscience, geoscience, biology, ecology, and mathematics, will place special demands on the educational system. It will require training of a new generation of researchers with special experimental, communication, and quantitative reasoning skills. Because the task of ecosystem analysis is too large to be tackled in an individual project, it may be necessary to reconsider the structure of graduate student training programs. It is possible that traditional, carefully delineated, individual PhD projects will be replaced by carefully integrated, collaborative PhD research efforts that include individuals at all career levels. Changes such as this will have the added advantage of generating scientists that are able to work together to solve large, complicated problems.

The integration of science and technology to develop understanding of the environment should extend to all educational levels. For example, an effective use of nanotechnology may be to monitor processes in the vicinity of K-12 classrooms (e.g., bird migrations, air quality, pesticide degradation in soil) and to compare these data to those collected elsewhere. This may improve the public's appreciation of the Earth's environments as complex biogeochemical systems that change in definable and predictable ways as the result of human activities.

Conclusions

Molecularly resolved analyses of environmental systems will allow us to determine how increasingly complex systems, from the level of cells and microbial communities up to entire ecosystems at the planetary scale, respond to environmental perturbations. With this knowledge in hand, we can move toward rigorous determination of environmental state and prediction of ecosystem change.

High-resolution molecular- and nanometer-scale information from both inorganic and biological components of natural systems will dramatically enhance our ability to utilize microbial processes (such as light-harvesting molecules for solar cells or mineral-solubilizing enzymes for materials processing) for technological purposes. This may be of great importance if we are to reduce our dependence on energetically expensive manufacturing and existing energy resources. For example, bioleaching is an alternative to smelting, bioextraction is an alternative to electrochemistry, biosynthesis of polymers is an alternative to petroleum processing, biomineralization is an alternative to machine-based manufacturing. Ultimately, nano-bio-geo integration will allow us to tease apart the complex interdependencies between organisms and their surroundings so that we may ultimately gain sufficient understanding of environmental systems to avoid the fate of microorganisms grown in a petri dish (Figure D.2).

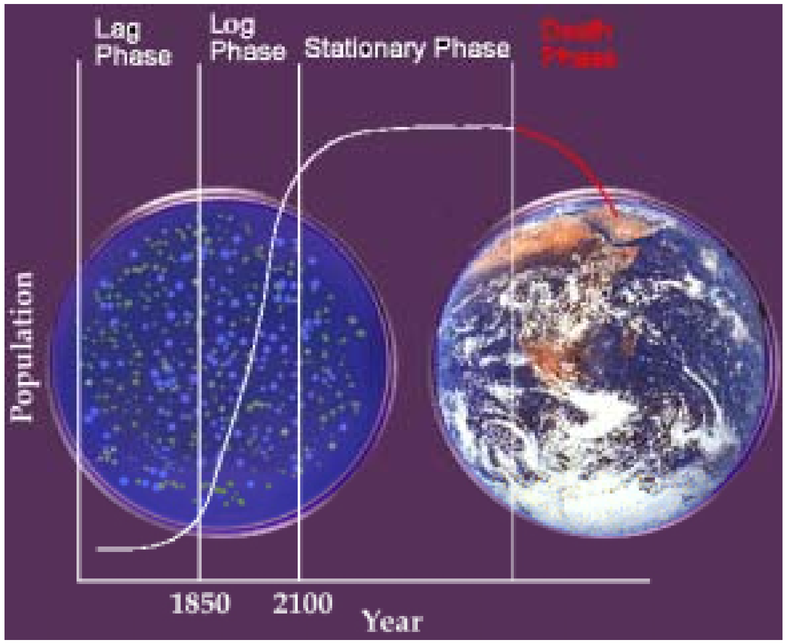


Figure D.2. Microbial communities growing within a confined space (here shown in a petri dish, left) have a cautionary tale to tell: overuse and/or unbalanced use of resources leads to build up of toxins, shortage of food, overpopulation, and death.

VISIONARY PROJECTS

THE COMMUNICATOR: ENHANCEMENT OF GROUP COMMUNICATION, EFFICIENCY, AND CREATIVITY

Philip Rubin, Murray Hirschbein, Tina Masciangioli, Tom Miller, Cherry Murray, R.L. Norwood, and John Sargent

As envisioned, *The Communicator* will be a “smart,” multifaceted, technical support system that relies on the development of convergent technologies to help enhance human group communication in a wide variety of situations, including meetings (both formal and informal), social exchanges, workplace collaborations, real-world corporate or battle training situations, and educational settings. This system will rely on expected advances in nanotechnology, fabrication, and a number of emerging information technologies, both software and hardware. In this system, these technologies will be tightly coupled with knowledge obtained from the biological and cognitive domains. The convergence of these technologies will serve to enhance existing attributes of individuals and remove barriers to group communication. This system will consist of a set of expanding implementations of these convergent technologies, growing more complex as the individual technologies mature over time. Some of these implementations are described below.

The initial goal of The Communicator is simple: to remove the kinds of barriers that are presently common at meetings where participants rely for communication on similar but slightly varying technologies. For example, it is standard for meeting participants to use software such as *PowerPoint* to present their ideas, but they often encounter technical difficulties moving between computers and computer platforms different from those on which they created their presentations. The transfer of information between systems during meetings is often hampered by varying media, connector differences, and incompatible data standards. At its simplest level, The Communicator would serve as an equalizer for communication in such situations, detecting the technological requirements of each participant and automatically resolving any differences in the presentation systems. The transfer and presentation of information would then become transparent.

Moving beyond this initial implementation, The Communicator would serve to remove more significant communication barriers, such as those related to physical disabilities or language differences. For example, the system, once apprised of a group member's hearing impairment, could tailor a presentation to that participant's needs by captioning the spoken or other auditory information. Similarly, it could produce auditory transcriptions of information presented visually in a group situation for any visually impaired member of the group. It could also provide simultaneous translation of meeting proceedings into a number of languages.

At the heart of The Communicator system are nano/info technologies that will allow individuals to carry with them electronically stored information about themselves that they can easily broadcast as needed in group situations. Such information might include details about preferences, interests, and background. Early implementations of this approach are doable now.

An even more interesting and advanced implementation would consist of detection and broadcast of the physiological and affective states of group participants with the purpose of providing resources to individuals and tailoring interactivity in order to allow the group to more easily achieve its goals. Detection of participants' physiological and affective states would be determined by monitoring biological information (such as galvanic skin response and heart rate) and cognitive factors via pattern recognition (such as face recognition to detect facial emotion, and voice pitch analysis to detect stress levels). Based on determinations of the needs and physical and cognitive states of participants, The Communicator could tailor the information it supplies to each individual, providing unique resources and improving productivity. Participants would have the ability to define or restrict the kinds of information about themselves that they would be willing to share with other members of the group.

As an example of this implementation, in an international conference or tribunal, each participant could select simultaneous translation of the discourse. Through PDA-like devices or biopatches, the system could measure the empathy levels or stress levels of all negotiators. A personal avatar would serve as a "coach" for each individual, recalling past statements, retrieving personal histories, and functioning as a research assistant to prepare material for use in arguments and deliberations. The system would facilitate the building of consensus by identifying areas of nominal disagreement and searching for common values and ideas.

Beyond facilitation of group communication. The Communicator could also serve as an educator or trainer, able to tailor its presentation and able to operate in a variety of modes, including peer-to-peer interaction and instructor/facilitator interaction with a group. The Communicator would function as an adaptive avatar, able to change its personal appearance, persona, and affective behavior to fit not only individuals or groups but also varying situations.

Design Components

Several important distinct components of The Communicator, introduced below, should be carefully designed to work together seemlessly.

The Individual Information Component

One key element of The Communicator system will be a model or record of each individual, including how each individual interacts with the environment and how s/he prefers to do so. This would include information like the language spoken, the preferred sensory channel, and limitations on input and output. This system should also include characteristics of the individual's cognitive capabilities: learning speed, preferences for learning modalities, areas of expertise, leisure activities, history of important social events, and other attributes that are relevant to a given task or situation. Ways that this element could be applied include the following:

- Using bioauthentication, the system could identify each individual in a group, including specific kinds of information about each individual. This could shorten the initial socialization process in a group setting.
- Users would be able to specify that they receive input translated into specific languages, including captioning or signing if needed.
- The system could determine what stress levels, information density, and learning rates work best for the individuals and the group as a whole.
- The system could provide support so that the individual learns in whatever modality works best for the topic at hand: auditory, haptic, text, images, virtual reality, and any specialized modality with which the user is comfortable. This would include such applications as using sounds to guide an individual through unknown territory if the person's vision and other senses are already monopolized by other inputs.

The Avatar Component

Another key element of The Communicator system will be avatars that can take on human appearance and behavior in a 3-D environment. They should be human-sized with full human fidelity, especially with respect to facial characteristics and emotion. The avatars should be able to assume any human form that is desired or most suitable (in terms of race, gender, and age, for example). The avatars' persona, mode of communication, and language should be able to be modified over time as the system learns the best method of communication or training for each individual.

The avatars should be life-like, so people will respond to them as though they are real. Avatars should be "in-a-box" and able to be placed and projected wherever needed, whether on a screen, as a hologram in the middle of a room, or through virtual reality viewers.

Possible applications include the following:

- Avatars could represent the human participants in a group to each other.
- They could also represent autonomous computerized agents that perform particular functions of the information and communication system.
- Avatars could be sent into dangerous situations, for example, to negotiate with a criminal holding a hostage.

- They could function as a resident nurse to the sick or as a companion to the elderly.
- An individual could perceive what his or her personal avatar encounters, e.g., “feeling” the presence of a biohazard or radiation in a dangerous environment while remaining immune to its harm.
- A training avatar (or a human tutor) could teach a person new skills by sharing the experience, for example, via a haptic suit, that could train a person in the physical movements required for dance, athletics, weaponry, or a refined manual skill such as surgery.

The Environmental Interface Component

A third key element of The Communicator system will be its interfaces with the surrounding “environmental network,” creating the opportunity for enhanced, personalized communications and education. Characteristics of how humans interact with information and technology can be viewed as constraints, or they can be viewed as strengths that convergent technology can play to. For example, if an individual is good at detecting anomalies or patterns in data, the technology would enhance this capability. Perhaps the technology would provide a “rheostat” of sorts to increase or decrease the contrast in data differences. This interface is a two-way street. The environment knows who is present, and each user receives appropriate information in the preferred form.

- The transforming strategy would apply known neural assessment techniques along with standard educational objectives, progressing to full cogno-assisted individualized learning in a group setting or collaborative learning.
- The system would be useful for teleconferencing, since participants need not be in the same location.
- The system should make it possible to adjust the social structure of communications, from whole-group mode in which all parties receive all messages, to more structured communication networks in which subgroups and individuals play specialized roles.

Key design considerations of The Communicator include the following:

- Very high-speed communications are needed, whether cable or wireless.
- The human-computer interface should be a wearable system offering augmented reality in office, schoolroom, factory, or field situations.

Educational Applications

All communication involves learning, but an *Educator* version of The Communicator could be created that would enhance many kinds of education. The convergence of NBIC technologies can radically transform the teaching and learning process and maximize the sensory and cognitive abilities of students. Some examples of applications include assistance to the learning disabled, optimally timed and individually presented learning experiences, and learning in a collaboratively orchestrated environment.

Several strategies could be employed to implement the Educator vision, geared for either individuals or groups, the classroom or the field. In the K-12 educational experience, a personal avatar or “coach” could govern hands-on experiments in accomplishing such goals as learning reading, science, math, or foreign languages. It would “teach” students as a human teacher does but would optimize itself to the needs of the student. It would be patient, friendly, stern, or take on any appropriate behavior. It might be most suitable for younger students but could also be a mentor for adults. If needed, it could be a "copilot." In a work environment, it could not only teach prepared lessons but also monitor performance and instruct on how to improve it.

The system could merge the following technologies:

- biotechnology to assess the physiological and psychological state of the learner, sense moods and states of mind
- cognitive science and technology to present responsive and individualized presentations of material to the student through different modalities
- expert information technology to accumulate and supply educational information

Military training could employ The Educator to teach decision-making under stress in a battlefield game in which the battlefield is virtual and the soldier is the general. As the war game is played, the avatar could be the general's assistant, read out after the battle. In another scenario, the virtual battlefield could be in a real field where soldier-participants wear wireless PDA helmets. The system could also be used as a "decision-making under stress" teaching tool for corporate executives.

Educator avatars could assume a wide variety of images (male, female, young, old) and be capable of speaking in all languages (oral and otherwise); identifying individual learning styles and then adapting curricula to individual needs; and using access to biological data to determine which methods are most effective for the assimilation and retention of knowledge. This could effectively improve education and training in all arenas from preschool through graduate school and across the corporate and military environments. It would equalize educational opportunities for all, enable learners to move through material at their own pace, and ensure that knowledge of learning styles would be retained and carried forward from year to year as children move from teacher to teacher or adults move from job to job.

Social Equalization

The adaptive capabilities of The Communicator would have the potential in group interactions to minimize the biases that arise from a variety of factors such as physical size and posture, gender, race, language, culture, educational background, voice tone and volume, and physical ability or disability. The result would be to maximize both individual and group performance. Examples include enhancing the performance of a poor learner, an athlete, or a soldier, and improving communication, collaboration, and productivity among people with a multitude of differences. Thus, the system would be not only a Communicator and Educator, but an *Equalizer* as well, enhancing human awareness, removing disabilities, empowering all members of society.

On a more f1undamental level, such a smart device could have a tremendous impact on the most disadvantaged people around the world, those who lack clean drinking water, adequate food supplies, and so on. Despite the lack of physical infrastructure like telephone cables, wireless Communicator technology could offer them the world of information in a form they can immediately use. Such knowledge will improve their agricultural production, health, nutrition, and economic status. No longer isolated from the global economic and cultural system, they will become full and valued participants.

Convergence

The Communicator system will incorporate each of the four NBIC technologies:

- *Nanotechnology* will be required to produce high-speed computational capabilities, wearable components that consume little energy, and pervasive sensors.

- *Biotechnology* will be fundamental to the interfaces, to monitoring the physical status of participants, and to the general design of human-friendly technologies.
- *Information technology* will be responsible for data management and transmission, translation across modalities and languages, and development of avatars and intelligent agents.
- *Cognitive science* will provide the understanding of effective learning styles, methods for elimination of biases, and the directions in which to search for common values and ideas that will be the foundation of a new form of social cooperation.

Some elements of The Communicator can be created today, but the full system will require aggressive research across all four of the convergent NBIC fields. Implementation of the entire vision will require an effort spanning one or two decades, but the payoff will be nothing less than increased prosperity, creativity, and social harmony.

ENHANCED KNOWLEDGE-BASED HUMAN ORGANIZATION AND SOCIAL CHANGE

Kathleen M. Carley, Carnegie Mellon University

Changes that bring together nanotechnology, information science, biology, and cognition have the potential to revolutionize the way we work and organize society. A large number of outcomes are possible. At the same time, existing social forms, legislation and culture will limit and direct the potential outcomes. In a very real sense, technologies and societies, tools and cultures, capabilities and legislation will co-evolve. Without attempting to predict the future, a series of possible outcomes, issues, and research challenges are discussed. Particular emphasis is placed on issues of security and potentially radical change within groups, organizations, and society.

Data and Privacy

In the area of bioterrorism, a key issue is early detection or “biosurveillance.” Early detection requires smart sensors at the biological level in the air, water, and ground, and on humans. Early detection requires integrating this data with geographic, demographic, and social information. Even were the sensors to exist, there would still be a problem: Under current legislation and privacy laws, the data cannot be integrated and made readily accessible to practitioners and researchers. To develop and test data mining tools, knowledge management tools, and what-if policy simulators, access is needed to a wide range of data in real time; but, providing access to such data enables the users of these tools to “know” details of individual behavior.

In the area of organizations, a key issue is team design and redesign (Samuelson 2000). Team design and redesign requires accurate data of who knows what, can work with whom, and is currently doing what. Doing such a skill audit, network analysis, and task audit is a daunting task. Maintaining the information is even more daunting. Individuals are loathe to provide the information for fear of losing their basis of power or anonymity, or for fear of reprisal. However, much of the information is implicit in the locations that people occupy, their stress levels, webpages, curricula vitae, public conversations, and so on.

In the cases of both acquiring and maintaining individual data, nano-bio-sensors that are embedded in the body and that report on individual health, stress level, and location; intelligent surfaces that track who is present while reshaping themselves to meet the needs of and enhance the comfort of the users; auto-sensors that create a memory of what is said, when people cough or sneeze; air and water sensors that sense contaminants; data-mining tools that locate information, simulation tools that estimate the change in social outcomes; information assurance tools and secure distributed databases all can be used to enable better outcomes. Indeed, such tools are critical to the collection, analysis, protection, and use of information to enhance group performance. The relatively easy problems here will be those that are dominated by technology, e.g., distributed database tools, data integration procedures, information assurance technology, and smart sensors. Those problems dealing with the need to change cultures, legislation, and ways of working will be more difficult. Privacy laws, for example, could mitigate the effectiveness of these tools or even determine whether they are ever developed. There are many critical privacy issues, many of which are well identified in the NRC report, *The Digital Dilemma* (http://www.nap.edu/catalog/9601.html). Views of knowledge as power will limit and impede data collection. Having such data will revolutionize healthcare, human resources, career services, intelligence services, and law enforcement. Having such data will enable "big-brotherism."

Were we able to overcome these two mitigating factors, then a key issue will become, "What will the bases for power be when knowledge is no longer a controlled commodity?" Since many organizations are coordinated and managed through the coordination and management of information, as knowledge is no longer controlled, new organizational forms should emerge. For example, a possible result might be the development of monolith corporations with cells of individuals who can do tasks, and as those tasks move from corporation to corporation, the cells would move as well. In this case, benefits, pay scales, etc., would be set outside the bounds of a traditional corporation. In this case, individual loyalty would be to the area of expertise, the profession, and not the company. Corporations would become clearinghouses linking agents to problems as new clients come with new problems.

Ubiquitous Computing and Knowledge Access

As computers are embedded in all devices, from pens to microwaves to walls, the spaces around us will become intelligent (Nixon, Lacey, and Dobson 1999; Thomas and Gellersen 2000). Intelligent spaces are generally characterized by the potential for ubiquitous access to information, people, and artificial agents, and the provision of information among potentially unbounded networks of agents (Kurzweil 1988). The general claim is that ubiquitous computing will enable everyone to have access to all information all the time. In such an environment, it is assumed that inequities will decrease. This is unlikely. While ubiquitous computing will enable more people to access more information more of the time, there will still be, short of major reforms, people with little to no access to computing. There will be excess information available, information making it difficult to discern true from false information. There will be barriers in access to information based on legislation, learning, and organizational boundaries. While information will diffuse faster, the likelihood of consensus being reached and being accurate given the information will depend on a variety of other factors such as group size, the complexity of the task and associated knowledge, initial distribution of information in the group, and do on. As a result, things may move faster, but not necessarily better.

Initial simulation results suggest that even when there are advanced IT capabilities, there will still be pockets of ignorance, certain classes of individuals will have privileged access to information and the benefits and power that derive from that, groups will need to share less information to be as or more effective, databases may decrease shared knowledge and guarantee information loss, and smaller groups will be able to perform as well or better than larger groups (Alstyne, M. v., and Brynjolfsson, E. 1996; Carley 1999). To address issues such as these, researchers are beginning to use multiagent network models. These models draw on research on social and organizational networks (Nohira and Eccles 1992), advances in network methodology (Wasserman and Faust 1994), and complex system models such as multiagent systems (Lomi and Larsen 2001). In these models, the agents are constrained and enabled by their position in the social, organizational, and knowledge networks. These networks influence who interacts with whom. As the agents interact, they learn, which in turn changes with whom they interact. The underlying networks are thus dynamic. The results suggest that organizations of the future might be flatter, with individuals coming and going from teams based on skills, that is, what they know, and not whom they know. As a result, social life will become more divorced from organizational life. Initial simulation results suggest that if information moves fast enough, decisions will become based not as much on information as on the beliefs of others; this should be particularly true of strategic decisions.

Socially Intelligent Technology

Major improvements in the ability of artificial agents to deal with humans and to emulate humans will require those artifacts to be socially intelligent. Socially intelligent agents could serve as intelligent tutors, nannies, personal shoppers, etc. Sets of socially intelligent agents could be used to emulate human groups/organizations to determine the relative efficacy, feasibility, and impact of new technologies, legislation, change in policies, or organizational strategy. At issue are questions of how social these agents need to be and what is the basis for socialness. It is relatively easy to create artificial agents that are more capable than a human for a specific well-understood task. It is relatively easy to create artificial agents that can, in a limited domain, act like humans. But these factors do not make the agents generally socially intelligent. One of the research challenges will be for computer scientists and social scientists to work together to develop artificial social agents. Such agents should be social at both the cognitive and precognitive (bio) level. Current approaches here are software-limited. They are also potentially limited by data; nanotechnology, which will enable higher levels of storage and processing, will also be necessary. That is, creating large numbers of cognitively and socially realistic agents is technically unfeasible using a single current machine. Yet, such agents need to exist on a single machine if we are to use such tools to help individuals manage change.

A key component of socialness is the ability to operate in a multiagent environment (Epstein and Axtell 1997; Weiss 1999). However, not all multiagent systems are composed of socially intelligent agents. For a machine to be socially intelligent, it needs to be able to have a “mental” model of others, a rich and detailed knowledge of realtime interaction, goals, history, and culture (Carley and Newell 1994). Socially intelligent agents need transactive memory, i.e., knowledge of who knows whom (the social network), who knows what (the knowledge network), and who is doing what (the assignment network). Of course this memory need not be accurate. For agents, part of the “socialness” also comes from being limited cognitively. That is, omniscient agents have no need to be social, whereas, as agents become limited — boundedly rational, emotional, and with a specific cognitive architecture — they become more social.

One of the key challenges in designing machines that could have such capabilities is determining whether such machines are more or less effective if they make errors like humans do. What aspects of the constraints on human cognition, such as the way humans respond to interrupts, the impact of emotions on performance, and so on, are critical to acquiring and acting on social knowledge? While we often see constraints on human cognition as limitations, it may be that socialness itself derives from these limitations and that such socialness has coordinative and knowledge benefits that transcend the limitations. In this case, apparent limits in individuals could actually lead to a group being more effective than it would be if it were composed of more perfected individual agents (Carley and Newell 1994).

A second key challenge is rapid development. Computational architectures are needed that support the rapid development of societies of socially intelligent agents. Current multiagent platforms are not sufficient, as they often assume large numbers of cognitively simple agents operating in a physical grid space as opposed to complex intelligent, adaptive, learning agents with vast quantities of social knowledge operating in social networks, organizations, and social space. Moreover, such platforms need to be extended to enable the co-evolution of social intelligence at the individual, group, and organizational level at differing rates and accounting for standard human processes such as birth, death, turnover, and migration.

A third challenge is integrating such systems, possibly in real time, with the vast quantities of data available for validating and calibrating these models. For example, how can cities of socially intelligent agents be created that are demographically accurate, given census data?

Socially Engineered Intelligent Computer Anti-Viruses and DDOS Defenses

Computer viruses have caused significant financial losses to organizations (CSI 2000). Even though most organizations have installed anti-virus software in their computers, a majority of them still experience infections (ICSA 2000). Most anti-virus software can not detect a new virus unless it is patched with a new virus definition file. New virus countermeasures have to be disseminated once a new virus is discovered. Studies of viruses demonstrate that the network topology and the site of the initial infection are critical in determining the impact of the virus (Kephart 1994; Wang 2000; Pastor-Satorras 2001). What is needed is a new approach to this problem. Such an approach may be made possible through the use of socially intelligent autonomous agents.

The Web and the router backbone can be thought of as an ecological system. In this system, viruses prey on the unsuspecting, and distributed denial of service attacks (DDOS) spread through the networks “eating” or “maiming” their prey. Viruses are, in a sense, a form of artificial life (Spafford 1994). One approach to these attacks is to propagate another “species” that can in turn attack these attackers or determine where to place defenses. Consider a computer anti-virus. Computer anti-viruses should spread fixes and safety nets, be able to “eat” the bad viruses and restore the machines and data to various computers, without, necessarily, the user’s knowledge. Such anti-viruses would be more effective if they were intelligent and able to adapt as the viruses they were combating adapted. Such anti-viruses would be still more effective if they were socially intelligent and used knowledge about how people and organizations use computers and who talks to whom in order to assess which sites to infiltrate when. We can think of such anti-viruses as autonomous agents that are benign in intent and socially intelligent.

Social Engineering

Combined nano-, bio-, info-, and cogno-technologies make it possible to collect, maintain, and analyze larger quantities of data. This will make it possible to socially engineer teams and groups to meet the demands of new tasks, missions, etc. The issue is not that we will be able to pick the right combination of people to do a task; rather, it is that we will be able to pick the right combination of humans, webbots, robots, and other intelligent agents, the right coordination scheme and authority scheme, the right task assignment, and so on, to do the task while meeting particular goals such as communication silence or helping personnel stay active and engaged. Social engineering is, of course, broader than just teams and organizations. One can imagine these new technologies enabling better online dating services, 24/7 town halls, and digital classrooms tailored to each student’s educational and social developmental level.

The new combined technologies are making possible new environments such as smart planes, “living” space stations, and so on. How will work, education, and play be organized in these new environments? The organizational forms of today are not adequate. Computational organization theory has shown that how groups are organized to achieve high performance depends on the tasks, the resources, the IT, and the types of agents. You simply do not coordinate a group of humans in a board room in the same way that you would coordinate a group of humans and robots in a living space station, or a group of humans who can have embedded devices to enhance their memory or vision.

Conclusion

These areas are not the only areas of promise made possible by combining nano-, bio-, info-, and cogno-technologies. To make these and other areas of promise turn into areas of advancement, more interdisciplinary research and training is needed. In particular, for the areas listed here, joint training is needed in computer science, organizational science, and social networks.

References

Alstyne, M. v., and E. Brynjolfsson. 1996. Wider access and narrower focus: Could the Internet Balkanize science? *Science* 274(5292):1479-1480.

Carley, K.M. forthcoming, Smart agents and organizations of the future. In *The handbook of new media*, ed. L. Lievrouw and S. Livingstone.

_____. forthcoming. Computational organization science: A new frontier. In Proceedings, Arthur M. Sackler Colloquium Series on Adaptive Agents, Intelligence and Emergent Human Organization: Capturing Complexity through Agent-Based Modeling, October 4-6, 2001; Irvine, CA: National Academy of Sciences Press.

_____. forthcoming, Intra-Organizational Computation and Complexity. In *Companion to Organizations,* ed. J.A.C. Baum. Blackwell Publishers.

Carley, K.M., and V. Hill. 2001. Structural change and learning within organizations. In *Dynamics of organizations: Computational modeling and organizational theories,* ed. A. Lomi and E.R. Larsen. MIT Press/AAAI Press/Live Oak.

Carley, K.M. 1999. Organizational change and the digital economy: A computational organization science perspective. In *Understanding the Digital Economy: Data, Tools, Research*, ed. E. Brynjolfsson, and B. Kahin. Cambridge, MA: MIT Press.

Carley, K.M., and A. Newell. 1994. The nature of the social agent. *J. of Mathematical Sociology* 19(4): 221-262.

CSI. 2000. CSI/FBI computer crime and security survey. *Computer Security Issues and Trends.*

Epstein, J., and R. Axtell. 1997. *Growing artificial societies.* Boston, MA: MIT Press.

ICSA. 2000. ICSA Labs 6th Annual Computer Virus Prevalence Survey 2000. ICSA.net.

Kephart, J.O. 1994. How topology affects population dynamics. In *Artificial life III,* ed. *C.G.* Langton. Reading, MA: Addison-Wesley.

Kurzweil, R. 1988. *The age of intelligent machines*. Cambridge, MA: MIT Press.

Lomi, A., and E.R. Larsen, eds. 2001. *Dynamics of organizations: Computational modeling and organizational theories*. MIT Press/AAAI Press/Live Oak.

Nixon, P., G. Lacey, and S. Dobson, eds. 1999. Managing interactions in smart environments. In Proceedings, 1st International Workshop on Managing Interactions in Smart Environments (MANSE '99), Dublin, Ireland, December 1999.

Nohira N. and R. Eccles, eds. 1992. *Organizations and networks: Theory and practice*. Cambridge, MA: Harvard Business School Press.

Pastor-Satorras, R., and A. Vespignani. 2001. Epidemic dynamics and endemic states in complex networks. Barcelona, Spain: Universitat Politecnica de Catalunya.

Samuelson, D. 2000. Designing organizations. *OR/MS Today*. December: 1-4. See also http://www.lionhrtpub.com/orms/orms-12-00/samuelson.html.

Spafford, E.H. 1994. Computer viruses as artificial life. *Journal of Artificial Life*.

Thomas, P. and H.-W. Gellersen, eds. 2000. Proceedings of the International Symposium on Handheld and Ubiquitous Computing: Second International Symposium, HUC 2000, Bristol, UK, September 25-27, 2000 .

Wang, C., J.C. Knight, and M.C. Elder. 2000. On computer viral infection and the effect of immunization. In Proceedings, IEEE 16th Annual Computer Security Applications Conference.

Wasserman, S. and K. Faust. 1994 *Social Network Analysis.* New York: Cambridge University.

Weiss, G., ed. 1999. *Distributed artificial intelligenc*e. Cambridge, MA: MIT Press.

A VISION FOR THE AIRCRAFT OF THE 21ST CENTURY

S. Venneri, M. Hirschbein, M. Dastoor, National Aeronautics and Space Administration

The airplane will soon be 100 years old. Over that period of time, it has evolved from the cloth and wood biplanes of the 1920s to the first all-metal single-wing aircraft of the 1930s, to the 100-passenger commercial transports of the 1950s, to the modern jet aircraft capable of reaching any point in the world in a single day. Nevertheless, the design of the modern airplane really has not changed much in the last fifty years. The grandfather of the Boeing 777 was the Boeing B-47 bomber designed in the late 1940s. It had a sleek, tubular aluminum fuselage, multiple engines slung under swept wings, a vertical tail, and horizontal stabilizers. Today, the fuselage is lighter and stronger, the wings more aerodynamic, and the engines much more efficient, but the design is a recognizable descendent of the earlier bomber.

The aircraft of the 21st century may look fundamentally different (Figure D.3). NASA is beginning to look to birds as an inspiration for the next generation of aircraft — not as a “blueprint,” but as a biomimetic mode (Figure D.4). Birds have evolved over the ages to be totally at home in the air. Consider our national bird, the eagle. The eagle has fully integrated aerodynamic and propulsion systems. It can morph and rotate its wings in three dimensions and has the ability to control the air flow over its wings by moving the feathers on its wingtips. Its wings and body are integrated for exceptional strength and light weight. And the wings, body, and tail work in perfect harmony to control aerodynamic lift and thrust and balance it against the force of gravity. The eagle can instantly adapt to variable loads and can see forward and downward without parallax. It has learned to anticipate the sudden drag force on its claws as it skims the water to grab a fish and how to stall its flight at just the right moment to delicately settle into a nest on the side of a cliff. The eagle is made from self-sensing and self-healing materials. Its skin, muscle, and organs have a nervous system that detects fatigue, injury, or damage, and signals the brain. The eagle will instantly adapt to avoid further trauma, and tissues immediately begin to self-repair. The eagle is designed to survive.

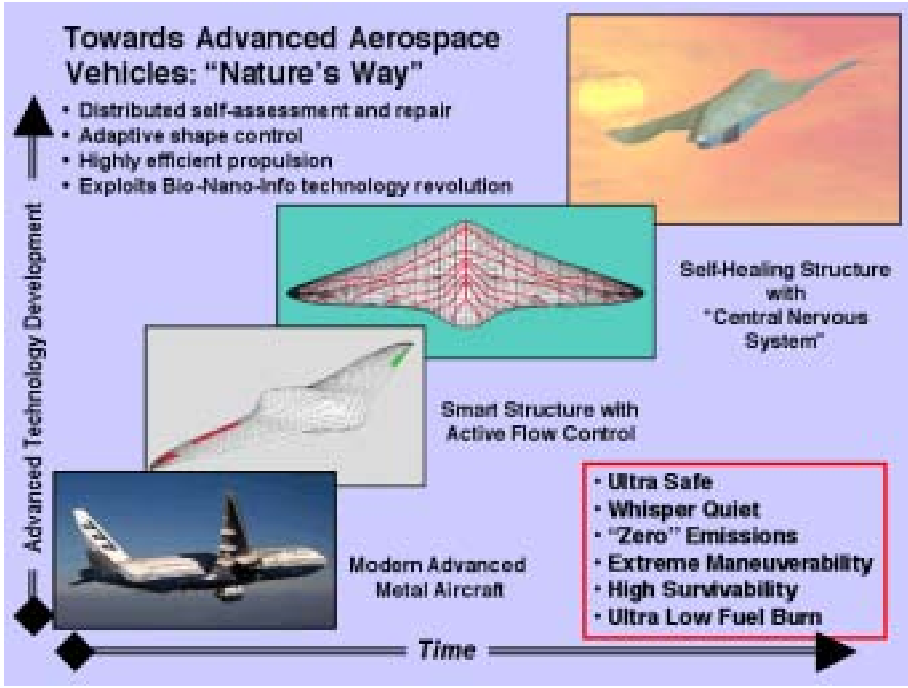


Figure D.3. Towards advanced aerospace vehicles: "Nature's Way."

Figure D.4. Inspiration for the next generation of aircraft.

NASA is pursuing technology today that is intended to lead toward just such a biomimetically inspired aircraft (Figure D.5). Advanced materials will make them lighter and more efficient to build. Advanced engines will make them fast and efficient. The airframe, engine, and cockpit will be "smarter." For decades, aircraft builders have worked to build wings that are stronger and stiffer. However, the wing that is needed for take-off and landing is not the wing needed for cruising. During take-off and landing, the wing needs to be highly curved from leading edge to trailing edge to produce enough lift at low speed. But this also produces a lot of drag. Once airborne, the wing needs to be flat for minimal drag during cruise. To change the wing shape, NASA has employed leading-edge slats — an articulated "nose" that runs along the length of the wing — and multipiece flaps that can drop the trailing edge of the wing by 60 degrees. All of this requires gear, motors, and hydraulic pumps.

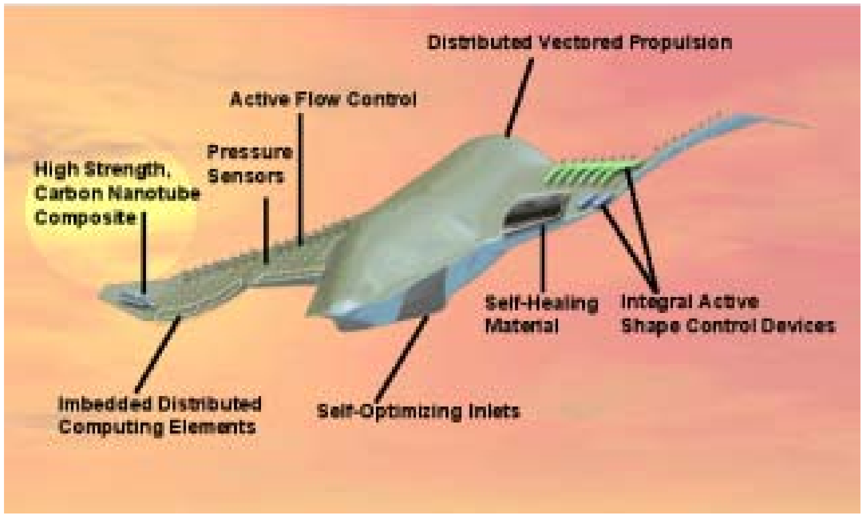


Figure D.5. NASA's dream of a future flight vehicle.

Imagine a bird-like wing of the future. It is not built from multiple, mechanically connected parts. It is made from new smart materials that have imbedded sensors and actuators — like nerves and sinew. The sensors measure the pressure over the entire surface of the wing and signal the actuators how to respond. But even the sensors are smart. Tiny computing elements detect how the aircraft responds to sensor signals. They eventually learn how to change the shape of the wing for optimal flying conditions. They also detect when there is damage to a wing and relay the extent and location to the pilot. And, like an injured bird, the wing adjusts its response to avoid further damage. This will not only be a very efficient and maneuverable airplane, but a very safe one.

Like the wings, the engines of this plane have integral health-management systems. Temperatures, pressures, and vibrations are all continuously monitored and analyzed. Unique performance characteristics are automatically developed for each engine, which then continually operates as efficiently as possible, and very safely. Long before a part fails, damage is detected and protective maintenance scheduled.

Inside the cockpit compartment, the pilot sees everything on a 3-D display that shows local weather, accentuates obstacles, all near-by aircraft, and the safest flight path. The on-board clear air turbulence sensor uses lasers to detect unsteady air well ahead of the aircraft to assure a smooth ride. When approaching a major airport, the lingering vortices that were shed from the wingtips of larger aircraft and that can upset a smaller one, can be easily avoided. This is a long-term vision, but emerging technology can make it real.

A key to achieving this vision is a fusion of nanoscale technology with biology and information technology (Figure D.6). An example is intelligent multifunctional material systems consisting of a number of layers, each used for a different purpose. The outer layer would be selected to be tough and durable to withstand the harsh space environment, with an embedded network of sensors, electrical carriers, and actuators to measure temperature, pressure, and radiation, and to trigger a response whenever needed. The network would be intelligent. It would automatically reconfigure itself to bypass damaged components and compensate for any loss of capability. The next layer could be an electrostrictive or piezoelectric membrane that works like muscle tissue with a network of nerves to stimulate the appropriate strands and provide power to them. The base layer might be made of biomolecular material that senses penetrations and tears and flows into any gaps. It would trigger a reaction in the damaged layers and initiate a self-healing process.

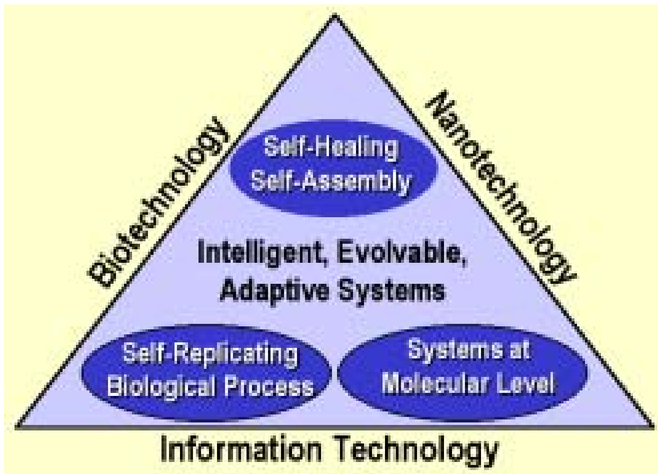


Figure D.6. Revolutionary technology vision as applied to future aircraft.

Carbon nanotube-based materials are an example of one emerging technology with the potential to help make this a reality. They are about a hundred times stronger than steel but one-sixth the weight of steel. They can have thermal conductivities seven times higher than the thermal conductivity of copper with 10,000 times greater electrical conductivity. Carbon nanotube materials may also have piezoelectrical properties suitable for very high-force activators. Preliminary NASA studies indicate that the dry weight of a large commercial transport could be reduced by about half compared to the best composite materials available today. The application of high-temperature nanoscale materials to aircraft engines may be equally dramatic. Through successful application of these advanced lightweight materials in combination with intelligent flow control and active cooling, thrust-to-weight ratio increases of up to 50 percent and fuel savings of 25 percent may be possible for conventional engines. Even greater improvement can be achieved by developing vehicle designs that fully exploit these materials. This could enable vehicles to smoothly change their aerodynamic shape without hinges or joints. Wings and fuselages could optimize their shape for their specific flight conditions (take-off, cruise, landing, transonic, and high-altitude).

In the long-term, the ability to create materials and structures that are biologically inspired provides a unique opportunity to produce new classes of self-assembling material systems without the need to machine or process materials. Some unique characteristics anticipated from biomimetics are hierarchical organization, adaptability, self healing/self-repair, and durability. In the very long term, comparable advances in electrical energy storage and generation technology, such as fuel cells, could completely change the manner in which we propel aircraft. Future aircraft might be powered entirely electrically. In one concept, thrust may be produced by a fan driven by highly efficient, compact electric motors powered by advanced hydrogen-oxygen fuel cells. However, several significant technological issues must still be resolved in order to use hydrogen as a fuel, such as efficient generation and storage of hydrogen fuel and an adequate infrastructure necessary for delivering the fuel to vehicles (Figure D.7).

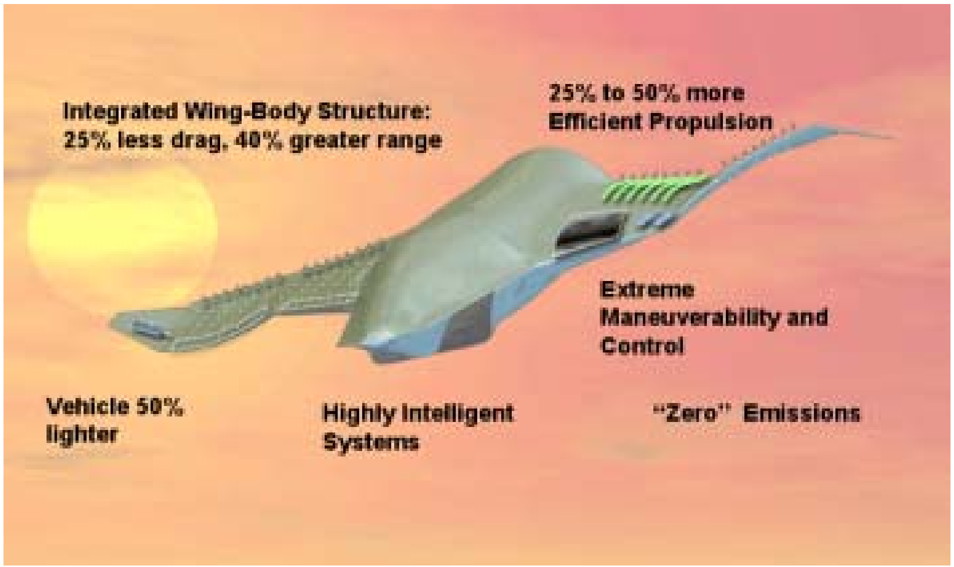


Figure D.7. Attributes of a future flight vehicle.

None of this is expected to happen quickly. Over the next decade we will likely see rapid development of advanced multifunctional, nanotechnology-based structural materials, such as carbon nanotube composites. Integrated health monitoring systems — for airframe and engine — may be developed, and deformable wings with imbedded actuators may also be developed. The cockpit will likely begin to become more of an extension of the pilot with greater use of senses other than sight to provide "situational awareness" of the aircraft and its operating environment. In two to three decades, we may see the first "bio/nano/thinking/sensing" vehicles with significant use of nanotechnology-based materials, fully integrated exterior-interior flow control, and continuously deformable wings. By then, the aircraft may also have a distributed control/information system — like a nervous system — for health monitoring, some level of self-repair, and cockpits that create a full sensory, immersive environment for the pilot.

MEMETICS: A POTENTIAL NEW SCIENCE

Gary W. Strong and William Sims Bainbridge, National Science Foundation [2]

In the "information society" of the twenty-first century, the most valuable resource will not be iron or oil but culture. However, the sciences of human culture have lacked a formal paradigm and a rigorous methodology. A fresh approach to culture, based on biological metaphors and information science methodologies, could vastly enhance the human and economic value of our cultural heritage and provide cognitive science with a host of new research tools. The fundamental concept is the *meme*, analogous to the gene in biological genetics, an element of culture that can be the basis of cultural variation, selection, and evolution.

The meme has been characterized both as a concept that could revolutionize the social sciences as the discovery of DNA and the genetic code did for biology, and as a concept that cannot produce a general

[2] The views in this essay do not necessarily represent the views of the National Science Foundation.

theory of social evolution because requirements for Darwinian evolution do not map into the social domain (Aunger 2000). There is a lot we do not understand about human behavior in groups, its relation to learning, cognition, or culture. There is no general theory that situates cognition or culture in an evolutionary framework, Darwinian or otherwise. It is also hard to conduct science in the social domain, not just because it is difficult to conduct experiments, but also because it is difficult to be objective. Prior efforts to "Darwinize" culture have a long and ignoble history. The question naturally arises as to what is new that might allow progress this time around, or should discretion take the better part of valor?

While any debate tends to sharpen the debate issues, in this case it may prematurely close off a search for a scientific definition of important terms and of an appropriate contextual theory. For example, a strictly Darwinian approach to cultural or social evolution may not be appropriate since humans can learn concepts and, in the same generation, pass them on to their offspring. Because memes are passed from one individual to another through learning, characteristics an individual acquires during life can be transmitted to descendents. This is one of the reasons why memes may evolve more rapidly than genes. In the language of historical debates in biology, culture appears to be Lamarckian, rather than Darwinian (Strong 1990). This would imply a different set of requirements for an evolutionary system that are not yet well understood.

As another example, we are only now discovering that many of the genes of an organism code for "chaperone" proteins that do not have "meaning" in a particular biological function, but, rather, play a role in molecular recycling and enabling the proteomic networks of molecules to interact in an orderly fashion (Kim et al. 1998). We do not yet understand how a balance is kept within a cell between the evolutionary need for variety and the need to preserve order in systems. Nevertheless, it is likely that in a fast-changing Lamarckian system, such processes become even more important. On the socio-cultural level, religious ideologies appear to have chaperone roles that may help keep individuals focused on important daily activities rather than getting caught up in unsolvable dilemmas and becoming unable to act. Even so, such ideologies cannot become so strict as to eliminate important variety from an evolutionary system. This tradeoff between order and disorder may operate like a regulator for social change (Rappaport 1988).

While there is no known Federal grants program focused on memetics, nor any apparent, organized research community, there are likely a number of existing and completed research projects that impact on the domain. These probably are found in a variety of disciplines and do not use a common vocabulary. For example, a few archaeologists apply evolutionary theory in their work (Tschauner 1994; Lyman and O'Brien 1998), and some cultural anthropologists explore the evolution of culture in a context that is both social and biological (Rindos 1985; Cashdan 2001; Henrich 2001). However, most archaeologists avoid theoretical explanations altogether, and cultural anthropology is currently dominated by a humanist rather than scientific paradigm. So, even though starting a research program in this area would not have to begin from scratch, there would be much work to do. The biggest roadblock would be getting researchers from various disciplines to collaborate over a common set of interests.

At a first approximation, there are three different realms in which biological genetics is valuable to humanity. First, it contributes to the progress of medicine, because there is a genetic aspect to all illnesses, not only to those diseases that are commonly labeled "genetic" or "inherited." Second, it provides valuable tools for agriculture, most recently including powerful techniques of genetic engineering to design plants and animals that are hardier, more nutritious, and economically more profitable. Third, it answers many of the fundamental scientific questions about the nature and origins of biological diversity, thus contributing to human intellectual understanding of ourselves and the world we live in. Cultural memetics would have three similar realms of applications, as described below.

Cultural Pathology

Culture is not just art, music, language, clothing styles, and ethnic foods. Importantly, it also includes the fundamental values, norms, and beliefs that define a society's way of life. Thus, the classic problem of social science has been to understand how and why some people and groups deviate from the standards of society, sometimes even resorting to crime and terrorism. Recent attention on closed groups has once again raised the question, "Why do people believe weird things?" — to borrow from a recent book title (Schermer 2002). The problem of social order thus depends upon the dynamic interactions between cultures, subcultures, and countercultures.

For decades, various anthropologists have considered whether or not there is a cultural equivalent of the human genome underlying differences of belief and behavior across groups or whether cultural context differentially expresses elements from a common repertoire available to all humans. One way to approach the issue might be to study culture with methodologies similar to those of bioinformatics.

A key bioinformatics construct is the genomic code, the cultural equivalent of which has been widely discussed under the concept of "meme" (Dawkins 1976). Cross-cultural signals are often undetected or misidentified, and cultural miscommunication is commonplace, leading one to suspect the existence of such codes and their differentiation among social groups. Levi-Strauss (1966) refers to cultural concepts, or artifacts, as "things to think with." Such shared concepts may, however, be more a form of externalized representation, or "cognitive Post-It Note," with important information processing functionality for a social group.

The prevalence of fundamentalist cultural and religious movements, for example, suggests that there may be an equivalent of the "auto-immune" response at the cultural level. Religion appears to be what Talcott Parsons (1964) called an "evolutionary universal," essential to the functioning of societies and prominent in every long-lasting culture. Within the realm of religion, diversification also appears to be universal, and it may be vain to hope that all people can eventually come to share compatible religious beliefs (Stark and Bainbridge 1987). At the present time, it is crucial to understand that revitalization or nativistic movements appear to be universal in times of great social change (Wallace 1956). Such movements tend toward increased orthodoxy and the involvement of charismatic leaders. Anthropologists have studied such movements from the time of the "Ghost-Dance" cults of native North Americans at the end of the 19th century to the rise of militant groups in Islam today (La Barre 1972).

"World-views" may be self-regulating, in this respect, each dominant ideology naturally stimulating the evolution of counter-ideologies. Just when Western Civilization rejoiced that it had vanquished Nazism and Marxism, and the "end of history" was at hand, radical Islam emerged to challenge its fundamental values (El-Affendi 1999). Quite apart from the issue of terrorist attacks from radical fringes of Islam, the entire Muslim religious tradition may have an evolutionary advantage over western secularism, because it encourages a higher birth rate (Keyfitz 1986). An inescapable natural law may be at work here, comparable to that which regulates the constantly evolving relations between predators and prey in the biological realm, ensuring that there is always a rival culture, and complete victory is impossible (Maynard Smith 1982). However, deep scientific understanding of the memetic processes that generate radical opposition movements may help government policymakers combat them effectively. It may never be possible to eradicate them entirely, but with new scientific methods, we should be able to prevent them from driving our civilization to extinction.

A science of memetics, created through the convergence of many existing disciplines, would likely give a basis for understanding the relationship between social groups and globalization — a topic of enormous recent interest. Fundamentalist groups are no longer "fringe" as they practice tactics to deal with variety and change, and they have become a topic not only for cultural anthropologists but also for law enforcement and governments in general. Certain “ideas” may have the force of a social virus that spreads as quickly and can have as deleterious effects on a population as do biological viruses (Boyd and Richerson 1985; Dennett 1995; Sagan 1997). It is important to examine such theories and to consider whether or not people are naturally vulnerable to “hacking” in the concept domain, as their computer networks are vulnerable in cyberspace. At the same time, memetics can help us understand the forces that promote cooperation between people and sustain culturally healthy societies (Axelrod 1990).

Memetic Engineering

Since long before the dawn of history, human beings have influenced the evolution of plants and animals, by domesticating them, breeding them, and now by engineering their genetic structure directly (Diamond 1997). Over the same span of millennia, humans became progressively more sophisticated in the processes by which they generate and transmit new culture, leading to the advanced electronic media of today. However, while agriculture in recent centuries has employed genetic science and technology of advancing complexity to domesticate plants and animals, the culture-based industries have not yet made use of memetic science.

It is important to realize that the term *culture* is defined very broadly by anthropologists and other social scientists. It is not limited to high artistic culture (symphonies, oil paintings, and great poetry), popular culture (rock music, best-selling novels, and dress styles), or intellectual culture (academic philosophies, schools of scholarship, and scientific theories). It also includes the practices of skilled professions, from surgery to litigation, financial accounting to bridge building, dentistry to uranium mining, and from auto mechanics to rocket science. The habitual patterns of behavior in families, neighborhoods, corporations, and government agencies are also forms of culture. We can say that culture refers to any pattern of thought and behavior that is shared through learning, rather than being rooted in biological inheritance.

We take for granted the assumption that government agencies like the National Science Foundation, National Institutes of Health, Defense Advanced Research Projects Agency, and Department of Energy should conduct fundamental scientific research that will ultimately be of benefit to manufacturing and transportation industries and to the military. At the same time, debates range over how heavily government should be involved in supporting culture through agencies like National Endowment for the Arts or National Endowment for the Humanities. But here we are discussing something very different from grants to support the work of artists and humanists. Rather, we refer to fundamental scientific research on the dynamics of culture, that will be of benefit to culture-creating and communication industries, and to national security through relations with other countries and through an improved ability to deal successfully with a wide range of nongovernmental organizations and movements.

If manufacturing creates the hardware of modern economies, the culture industries create the software. Both are essential to prosperity, and in the modern world, both should be grounded in solid scientific knowledge. If we understood better how human beings actually innovate, whether in music or the engineering design of consumer products, we could help them do it better. If we had a better map of culture, analogous to the Linnean system that classifies biological organisms into species and genera, we could help people find the culture they want and we could locate “uninhabited” cultural territories that could profitably be colonized by growing industries. Many of the social problems faced by contemporary American society seem to have substantial cultural aspects, so the findings of scientific memetics would be extremely valuable for both the government agencies and private organizations that have to deal with them.

As the Human Genome Project drew to its conclusion, it became clear to everyone that "mapping the human genome" was only part of the work. Also necessary was studying the great genetic diversity that exists from person to person around the planet, and discovering the biochemical pathways through which each gene was expressed in the phenotypic characteristics of the individual. Comparable work will be required in cultural memetics. For any given cultural trait, there may exist a number of distinct alternatives, like alleles in biological genetics, the mutational forms of a gene. The characteristics of varied individuals are the complex result of different alleles interacting across numerous genes. Categorization of culture from a memetic perspective will identify these alleles, and memetic engineering could make extensive use of techniques for combining these cultural traits in new ways (Bainbridge 1985).

Understanding how memes are expressed in actual human behavior will require advances in cognitive science that will have spin-off benefits in education and the culture industries. For example, research on how language is encoded both memetically and cognitively will contribute to better language instruction in schools and more effective commercial and governmental translation across languages. As in any major scientific endeavor, there may be a large number of unexpected benefits, but the gains we can identify now already more than justify the development of memetic science on economic grounds alone.

A Science of Culture

Participants in the Convergent Technologies (NBIC) conference recommended a new scientific initiative, analogous to the Human Genome Project that charted the human genetic code, which they called the Human Cognome Project — unraveling the secrets of the human cognitive genome. Any attempt to solve the riddles of the human mind will have to be far more than an exercise in brain neurology; most importantly, it will have to attack the mysteries of the cultural genome.

One major benefit of a program in memetics would be to better understand culture as an evolutionary process in its own context, whether as a Darwinian, Lamarckian, or as yet unknown system (Boyd and Richerson 1985). The knowledge gained could create a framework for a scientific rebirth in social and cultural domains. While opinions vary, it would not be too harsh to suggest that several of the social sciences seem to have stalled, some of them achieving very little progress in recent decades. The same thing occasionally happens in physical sciences. For example, planetary astronomy had practically stalled in the two or three decades prior to the launch of the first interplanetary space probes. Similarly, cancer research has achieved progress only very slowly over the past century, but the Human Genome Project offers new hope of breakthroughs. Memetic science could provide just the intellectual boost and potent research methodology needed by such diverse fields as Anthropology, Political Science, and Sociology.

Development of new theories and methods will require cooperation between hundreds of scientists in perhaps a dozen fields, so here with our limited perspectives we can suggest only a few of the possibilities. Perhaps there are a number of common features of natural codes, including both cultural and biological codes:

- *The "independence" feature*: Natural code elements tend to have arbitrary meaning (C.S. Peirce's symbols, as opposed to icons or indices) facilitating abstraction and reuse.

- *The "combinatorial advantage" feature*: The number of potential representations is much larger in combinations of elements than in one-to-one element coding — perhaps because evolutionary selection favors representational richness available by combination sets.

- *The self-regulation of natural codes*: Dependency upon a code results in a constraint for new input to be interpreted in terms of the code; change is thereby limited to evolution of the code over time.

Work on applying language modeling to genomic sequences at Carnegie Mellon University has suggested that genomes differentiate species by having distributions that include rare occurrences and where such rare occurrences can often be species-unique. This work suggests that some species-unique sequences have an unusual generative power, such as those playing an important role in fold initiation of proteins. Perhaps cultural codes also contain some rare occurrences that serve to differentiate cultures and are heavily associative, or generative, within the culture.

The study of cultural codes, such as suggested here, has not proceeded as rapidly as other fields such as bioinformatics. Perhaps there are reasons of politics and objectivity that have lowered the expectation of resources available for doing such research. Cultural codes may be easier and more politically feasible to study in the short-run in culturally primitive groups or other large-brained species. Bottlenose dolphins, for example, participate in fluid, short-term social associations, and their vocal plasticity as well as their behavior appears to be related to their fission/fusion social structure (Reiss et al. 1997). Perhaps dolphins' fluid social groups provide external cognitive representations (perhaps via "mirror neurons") in a manner similar to the totems of primitive human cultural groups.

Several systematic research methodologies need to be developed. One breakthrough that seems within reach would be the memetic equivalent of the Linnean system for classifying species, genera, and other kinds of biological clades. In recent years, information science has developed a range of techniques, such as latent semantic analysis and semantic concept space technology (Harum et al. 1996). United with cognitive science, these methods should go a long way to identifying the structure of the cultural genome and the mechanisms by which it changes or sustains itself. Through the development of memetic science, we will want to look to genetics for inspiration and selectively import both theories and methods from biology when appropriate.

The scientific study of culture is both possible and pregnant with knowledge of human behavior. Thus, it deserves to be given more resources, especially in light of current events. These events include not only the terrorism of September 11, 2001, but also the dot-com crash and the failure of nations as diverse as Argentina, Indonesia, and Japan to sustain their economic development. Memetic science could help us deal with challenges to American cultural supremacy, discover the products and services that will really make the information economy profitable, and identify the forms of social institutions most conducive to social and economic progress.

A Transforming Strategy

The most obvious barrier to the emergence of a successful science of memetics is the lack of a unified scientific community to create it. We suggest that three kinds of major projects would be needed to establish the nucleus for this vital new field:

1. *Professional conferences, scientific journals, and a formal organization devoted to memetics.* A scientific community needs communication. Because memetics spans biology, information science, cognitive science, and cultural studies, the people who will create it are strewn across many different disciplines that hold their annual meetings at different times in different cities. Thus, a series of workshops and conferences will be essential to bring these people together. Out of the conferences can emerge publications and other mechanisms of communication. An electronic communication network at the highest level of scientific quality needs to be established.

2. *Data infrastructure, in the form of multiuse, multiuser digital libraries incorporating systematic data about cultural variation, along with software tools for conducting scientific research on it.* Some work has already been accomplished of this kind, notably the decades-long efforts to index the findings of cultural anthropological studies of the peoples of the world, accessible through World Cultures Journal (http://eclectic.ss.uci.edu/~drwhite/worldcul/world.htm), and cross-cultural questionnaire surveys such as The World Values Survey (http://wvs.isr.umich.edu/). However, existing data were not collected with memetic analysis in mind. They typically ignore most dimensions of modern cultures, and they lack information about the networks of communication between individuals and groups that are fundamental to memetic mutation and diffusion. Thus, entirely new kinds of cultural data infrastructure are needed, to provide the raw material for memetic science.

3. *Specific major research projects assembling multidisciplinary teams to study distinct cultural phenomena that are most likely to advance fundamental memetic science and to have substantial benefits for human beings.* Because culture is highly diverse, it is essential to support multiple projects in different domains. This strategy would connect data infrastructure projects with teams of scientists oriented toward answering specific but profound scientific questions. One recent suggestion that has merit on both scientific and practical grounds is to create an distributed digital library devoted to all aspects of Islamic culture, with special attention to understanding how it evolves and divides. Another worthwhile project would be to link existing linguistic data archives, for example represented by the Linguistic Data Consortium, then transform them into a laboratory for studying the constant process of change that goes on within and across languages. A very different project, with a wide range of intellectual and economic benefits, would be an institute to study the transformation of engineering and manufacturing by the development of nanotechnology, gaining fundamental scientific understanding of the innovation process, to improve the methods by which new technologies are developed.

References

Aunger, R., ed. 2000. *Darwinizing culture: The status of memetics as a science*. Oxford: Oxford University Press.

Axelrod, R. 1990. *The evolution of cooperation*. New York: Penguin.

Bainbridge, W.S. 1985. Cultural genetics. In *Religious movements*, ed. R. Stark. New York: Paragon.

Boyd, R. and P.J. Richerson. 1985. *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Dawkins, R. 1976. *The selfish gene*. Oxford: Oxford University Press.

Dennett, D.C. 1995. *Darwin's dangerous idea*. New York: Simon and Schuster.

Diamond, J. 1997. *Guns, germs, and steel: The fates of human societies*. New York: Norton.

El-Affendi, A. 1999. Islam and the future of dissent after the "end of history." *Futures* 31:191-204.

Harum, S.L., W.H. Mischo, and B.R. Schatz. 1996. Federating repositories of scientific literature: An update on the Digital Library Initiative at the University of Illinois at Urbana-Champaign. *D-Lib Magazine*, July/August, www.dlib.org.

Keyfitz, N. 1986. The family that does not reproduce itself. In *Below-replacement fertility in industrial societies: Causes, consequences, policies*, ed. K. Davis, M.S. Bernstam, and R. Campbell. (A supplement to *Population and Development Review*).

Kim, K.K., R. Kim, and S.-H. Kim. 1998. Crystal structure of a small heat-shock protein. *Nature* 394:595-599.

Levi-Strauss, C. 1966. *The savage mind*. Chicago: University of Chicago Press.

Lyman, R.L., and M.J. O'Brien. 1998. The goals of evolutionary archaeology: History and explanation. *Current Anthropology* 39:615-652.

Maynard Smith, J. 1982. *Evolution and the theory of games.* New York: Cambridge University Press.

Reiss, D., B. McCowan, and L. Marino. 1997. Communicative and other cognitive characteristics of bottlenose dolphins. *TICS* 140-145.

Strong, G.W. 1990. Neo-Lamarckism or the rediscovery of culture. *Behavioral and Brain Sciences* 13: 92.

Parsons, T. 1964. Evolutionary universals in society. *American Sociological Review* 29: 339-357.

Rappaport, R. 1988. *Ecology, meaning, and religion*. Richmond, California: North Atlantic Books.

Sagan, C. 1997. *The demon-haunted world: Science as a candle in the dark*. New York: Ballantine Books.

Schermer, M. 2002. *Why people believe weird things: Pseudoscience, superstition, and other confusions of our time*. New York: H. Holt.

Stark, R. and W. S. Bainbridge. 1996. *A theory of religion*. New Brunswick, NJ: Rutgers University Press.

Tschauner, H. 1994. Archaeological systematics and cultural evolution: Retrieving the honour of culture history. *Man* 29:77-93.

Wallace, A.F.C. 1956. Revitalization movements. *American Anthropologist* 58:264-281.